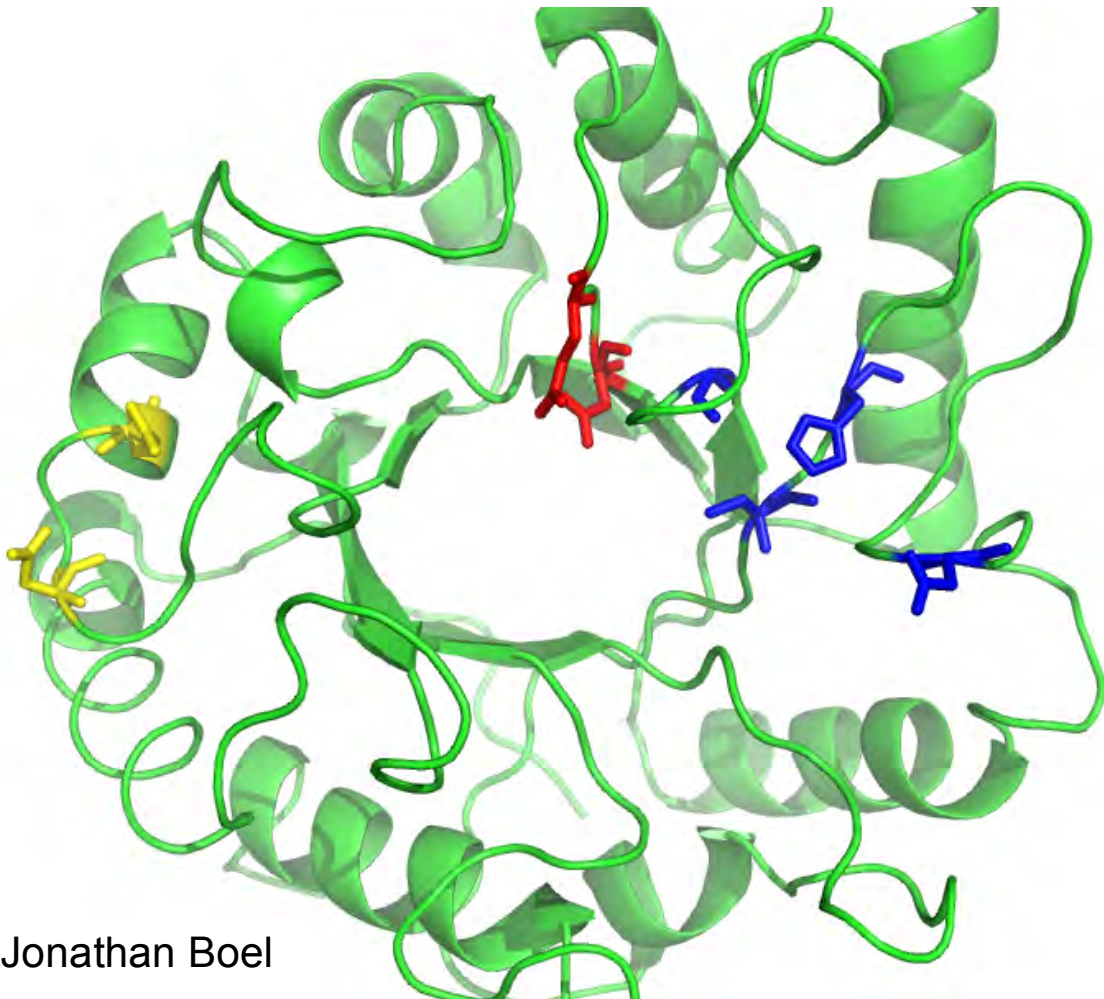


# Structure-Function Study on the Activation Mechanism of Endo T

Crystallization and biochemical characterization of the intact  
Endo T from *Trichoderma reesei* (*Hypocrea jecorina*)



Jonathan Boel

Department of Molecular Biology

Swedish University of Agricultural Sciences

Uppsala, 2011

EX0565, Independent project in biology, 30 hp, level E

## **Structure-Function Studies on the Activation Mechanism of Endo T**

Crystallization and biochemical characterization of the intact Endo T from *Trichoderma reesei* (*Hypocrea jecorina*)

### **Jonathan Boel**

Supervisors: Mats Sandgren<sup>1</sup>, Saeid Karkehabadi<sup>1</sup>, Stals Ingeborg<sup>2</sup>

<sup>1</sup>Dept of Molecular Biology, Swedish University of Agricultural Sciences (SLU) Sweden;

<sup>2</sup>Faculty of Applied Engineering Sciences, University College Ghent (Hogent), Ghent, Belgium;

Examiner: Jerry Ståhlberg

Department of Molecular Biology

SLU, Swedish University of Agricultural Sciences

Box 590, SE-751 24 Uppsala, Sweden



The MSc project was performed within the framework of the thematic research program “MicroDrive - Microbially Derived Energy” and the Erasmus program at the Faculty of Natural Resources and Agricultural Sciences at SLU and in collaboration with the Faculty of Applied Engineering Sciences, University College Ghent, Ghent, Belgium

## Index

List of figures .....	iv
List of tables .....	v
List of abbreviations .....	v
Abstract .....	1
1 Introduction .....	2
2 Literature .....	2
2.1 Expression organisms .....	2
2.1.1 <i>Trichoderma reesei</i> .....	2
2.1.2 <i>Pichia pastoris</i> .....	3
2.2 Protease .....	3
2.2.1 Proteases: an introduction.....	3
2.2.2 Protease activity in <i>T. reesei</i> .....	4
2.2.3 Protease activity in <i>P. pastoris</i> .....	4
2.3 Endo-N-acetyl- $\beta$ -D-glucosaminidase.....	5
2.3.1 EC classification .....	5
2.3.2 Glycoside Hydrolase family classification.....	5
2.3.3 Glycoside hydrolase family 18.....	7
2.4 N-Glycosylation .....	7
2.4.1 N-glycosylation an introduction .....	7
2.4.2 N-glycosylation in filamentous fungi.....	8
2.4.3 N-glycosylation in <i>Trichoderma reesei</i> .....	8
2.5 Endo T .....	8
2.5.1 Substrate .....	8
2.5.2 Mechanism .....	8
2.5.3 Structure.....	9
2.6 X-ray crystallography .....	9
2.6.1 Protein crystallography .....	9
2.6.2 X-ray diffraction .....	11
2.6.3 Electron density map.....	11
2.6.4 Protein structures .....	12
3 Materials and methods.....	13
3.1 Protein samples .....	13

3.2	Endo T protein analysis .....	13
3.2.1	SDS-PAGE analysis .....	13
3.2.2	Native-PAGE analysis .....	13
3.2.3	Peptide fingerprinting by mass spectrometry .....	13
3.2.4	Western Blot.....	13
3.2.5	Protein concentration .....	13
3.3	Purification of the intact Endo T for crystallization .....	14
3.3.1	Size-exclusion purification with a 3 kDa cutoff filter. ....	14
3.3.2	Superdex 75 size-exclusion chromatography .....	14
3.3.3	MonoQ anion exchange chromatography .....	14
3.3.4	Deglycosylation of intact Endo T for crystallization .....	14
3.4	Proteolytic stability of the intact Endo T .....	14
3.4.1	pH-stability.....	14
3.4.2	Temperature stability.....	15
3.4.3	Proteolytic degradation.....	15
3.4.1	Proteolytic activity.....	15
3.5	X-ray crystallography .....	15
3.5.1	Protein crystallization .....	15
3.5.2	X-ray diffraction .....	15
3.5.3	Protein structure and refinement .....	15
4	Results .....	16
4.1	Endo T protein analysis .....	16
4.1.1	Endo T transport control.....	16
4.1.2	Protein concentration .....	16
4.2	Purification of the intact Endo T for crystallization .....	16
4.2.1	Size exclusion chromatography .....	16
4.2.2	Anion exchange chromatography.....	16
4.2.3	Deglycosylation of intact Endo T for crystallization .....	17
4.3	Proteolytic stability of the intact Endo T .....	18
4.3.1	T- dependent degradation pattern of Endo T .....	18
4.3.2	pH-dependent degradation pattern of Endo T .....	19
4.3.3	Proteolytic degradation in crystallization .....	20
4.3.4	Proteolytic activity on Cel7A.....	20
4.4	Protein crystallization, data collection and structure refinement.....	21

4.5	Structural comparison.....	23
4.5.1	Comparison to known proteins.....	23
4.5.2	Comparison to unknown proteins.....	23
4.6	Possible protease active site .....	24
4.6.1	Serine protease .....	24
4.6.2	Acid protease .....	24
5	Discussion.....	25
5.1	Biochemical characterization.....	25
5.1.1	C-terminal stability.....	25
5.1.2	Deglycosylation .....	25
5.2	Protease activity .....	26
5.3	Endo T structure .....	26
5.3.1	N-terminus.....	26
5.3.2	C-terminus.....	26
5.1	Concluding remarks.....	27
6	Acknowledgments.....	28
7	References.....	29
8	Appendix .....	33
8.1	Appendix 1: Bio-Rad assay results.....	33
8.2	Appendix 2: Purified protein concentrations .....	33
8.3	Appendix 3: Western blot SDS-PAGE .....	34
8.4	Appendix 4: Structural comparison of known endo- $\beta$ -D-glucosaminidases.....	35
8.5	Appendix 5: Endo T eukaryotic blast results .....	36
8.6	Appendix 6: Genealogic tree from Endo T BLAST results .....	37
8.7	Appendix 7: Sequence alignment from Endo T BLAST results .....	37

## List of figures

Figure 1: ( $\beta/\alpha$ ) <sub>8</sub> TIM barrel structure from Endo T .....	7
Figure 2: Different N-glycosylation forms.....	7
Figure 3: Retaining mechanism with a N-acetyl or N-glycol group .....	9
Figure 4: A cartoon representation of the Endo T structure with a ligand bound in the active site.....	9
Figure 5: Protein crystallization phase diagram.....	9
Figure 6: Different vapour diffusion methods.....	10
Figure 7: Size exclusion chromatography of Endo T with protease inhibitors. ....	16
Figure 8: Protein status of Endo T after transport.....	16
Figure 9: Deglycosylation of inactive Endo T.....	17
Figure 10: Chromatogram of anion exchange chromatography with a MonoQ column .....	17
Figure 11: Different fractions of the first MonoQ purification. ....	17
Figure 12: The resulting gel after Native poly-acrylamide gel.....	17
Figure 13: SDS-PAGE showing the degradation pattern of Endo T incubated at 4°C .....	18
Figure 14: SDS-PAGE showing the degradation pattern of Endo T incubated at 21°C .....	18
Figure 15: SDS-PAGE showing the degradation pattern of Endo T incubated at 37°C .....	18
Figure 16: SDS-PAGE showing the degradation pattern of Endo T incubated at pH 3.....	19
Figure 17: SDS-PAGE showing the degradation pattern of Endo T incubated at pH 5.....	19
Figure 18: SDS-PAGE showing the degradation pattern of Endo T incubated at pH 8.....	19
Figure 19: SDS-PAGE showing the degradation pattern of Endo T incubated at pH 11.6.....	19
Figure 20: SDS-PAGE showing the degradation pattern of Endo T incubated at pH 3.....	19
Figure 21: SDS-PAGE showing the protein status in the crystals and in the crystal drops .....	20
Figure 22: SDS-PAGE of the Endo T protein crystallization samples. ....	20
Figure 23: The region in the protein sequence with the possible catalytic protease site.....	23
Figure 24: BLAST results.....	23
Figure 25: The possible serine protease active site positioned next to the ENGase active site.....	24
Figure 26: Cartoon representation of the pig trypsin protease catalytic site .....	24
Figure 27: Cartoon representation of Endo T superimposed on Endo F3 .....	25
Figure 28: Comparison of the N-terminus for the different Endo T structures at different pH. ....	26
Figure 29: Side view of the short C-terminus.....	27
Figure 30: Side top view of the short C-terminus.....	27
Figure 31: Top view of the short C-terminus.....	27
Figure 32: BSA standard curve for total protein concentration. ....	33
Figure 33: First SDS-PAGE for western blot. ....	34
Figure 34: Second SDS-PAGE for western blot. ....	34
Figure 35: Structural comparison of known endo- $\beta$ -D-glucosaminidases.....	35
Figure 36: Endo T eukaryotic blast results. ....	36
Figure 37: Genealogic tree from Endo T BLAST results. ....	37
Figure 38: Sequence alignment from Endo T BLAST results. ....	37

## List of tables

Table 1: Different proteases with their classification.....	3
Table 2: Endo-N-acetyl- $\beta$ -D-glucosaminidase in the different GH families.....	6
Table 3: Different crystallization conditions on which data was collected.....	21
Table 4: Data collection on the different crystals shown in table 3. ....	21
Table 5: Final refinement and model statistics. ....	22
Table 6: Ramachandran plot (coot) for the final refined Endo T structures. ....	22
Table 7: Unpurified Endo T total protein concentrations.....	33
Table 8: Purified Endo T protein concentrations.....	33

## List of abbreviations

AP.....	acid protease
Endo T.....	endo-N-acetyl- $\beta$ -D-glucosaminidase from <i>T.reesei</i> .
ENGase.....	endo-N-acetyl- $\beta$ -D-glucosaminidase
GH .....	Glycoside hydrolase family
Man.....	mannose
NAG.....	N-acetyl- $\beta$ -D-glucosamine
PDB.....	Protein Data Bank
SP.....	serine protease

## **Abstract**

Endo T is a deglycosylating enzyme secreted by the filamentous fungus *Trichoderma reesei* (*Hypocrea jecorina*). The active Endo T protein is both deglycosylated and C-terminal processed. Since the intact protein was never observed in the *T. reesei* culture medium, a construct of the intact Endo T was expressed in *P. pastoris*. Here, the glycosylated, intact and inactive protein was found in the medium but this form is slowly converted to the deglycosylated and proteolytic cleaved form. The structure of the active Endo T was already solved in previous work. We tried to solve the structure of the intact inactive Endo T to compare it with the active structure. Several structures of Endo T were solved, but none of them were structures of the intact Endo T. Different pH and T parameters were screened to find a condition where the C-terminal peptide was stable and the enzyme kept intact. A protease inhibitor cocktail was also added to limit the proteolytic process. We were able to control the proteolysis with the help of the protease inhibitor cocktail at pH 5 and higher, but we were not able to limit the proteolysis at pH 3, nor in any of the crystallization samples.

## **Keywords**

Endo T, glycoside hydrolase family 18, EC 3.2.1.96, deglycosylation, proteolysis, C-terminus, protein structure.



## 1 Introduction

Endo T is an endo-N-acetyl- $\beta$ -D-glucosaminidase (ENGase) secreted by the filamentous fungus *Trichoderma reesei*. It is an enzyme that hydrolyzes the  $\beta$ 1-4 bond between two acetyl glucosamine residues in high-mannose N-glycans. Glycosylation is an important post-translational modification involved in protein folding mechanisms and important for protein stability. ENGases are present in eukaryotes as part of the ERAD degradation pathway of misfolded proteins. Especially with prokaryotes ENGases are often involved in parasitism (Karamanos 1997). Endo T from *T. reesei* is responsible for the extensive deglycosylation observed for its cellulases and other secreted proteins, but its biological role is still unclear.

Endo T has been expressed both homologously in *Trichoderma reesei* and heterologously in *Pichia pastoris*. The *P. pastoris* expressed protein is secreted as an intact protein, fully glycosylated but inactive form, while the *T. reesei* expressed protein has lost 9 and 49 amino acids at its N- and C-terminus respectively and contains single N-acetyl glucosamine attached to Asn 70 and Asn 240 of the protein. This lower molecular weight enzyme form is catalytically active and hydrolyzes the  $\beta$ 1-4 bond in the chitobiose core of high mannose N-glycans. These can be bound on a glycoprotein or a glycopeptide, but the enzyme also acts on free N-glycans.

Endo T has EC classification 3.2.1.96 and is classified as a glycoside hydrolase (GH) family 18 enzyme. GH18 enzymes are retaining glycoside hydrolases using neighboring group participation with Asp and Glu as the catalytic residues. These two catalytic residues are conserved throughout the entire GH 18 family.

Previously the structure of the active Endo T expressed in *T. reesei*, both with and without ligand has been solved to a resolution of 1.3 Å. The aim for this master project was to solve the

structure of the intact and presumably inactive form of Endo T. This new structural knowledge about Endo T will help us to better understand how and why Endo T is activated. Solving the structure of the intact Endo T might be problematic. First off all it is possible that the C-terminal peptide and/or the N-glycans are too flexible to be visible in the electron density map for the protein. The second problem, originates from previous observations that the protein is very hard to keep intact over an extended period of time. The proteolytic cleavage of Endo T is difficult to prevent even with a purified protein sample. From the moment a little amount of active protein is formed, the deglycosylation process starts to proceed very fast. The reason why the C-terminus of the protein is removed so fast is not clear. The fast deglycosylation is due to Endo T that is deglycosylating itself.

## 2 Literature

### 2.1 Expression organisms

#### 2.1.1 *Trichoderma reesei*

*Trichoderma reesei* is a mesophile filamentous fungus from the division of the ascomycota. The subdivision is peizomycotina with sordariomycetes as class and hypocreaceae as family. *Trichoderma* is the genus name whereas *T. reesei* is the species name. *T. reesei* is an anamorphous or asexual reproduction form of *Hypocrea jecorina* (Taylor, Spatafora and Berbee 2006, Harman n.d.).

*T. reesei* is known to produce high amounts of cellulose-degrading enzymes. This is why modified strains of this fungus are commercially used for production of cellulases and other enzymes that degrade complex plant cell wall polysaccharides (Harman n.d., Sharma, et al. 2009). *T. reesei* is also of interest for its high production of extracellular heterologous proteins and is often used as a host organism (Maras, et al. 1997, Dienes, et al. 2007, Sharma, et al. 2009).

The *T. reesei* (hemi)-cellulases are partially of interest for 2<sup>nd</sup> generation bio-ethanol production. Instead of using starch or sucrose for the production of bio-ethanol it is possible now to use cellulose rich plant material as feedstock for the production of bio-ethanol (European biofuels TECHNOLOGY PLATFORM n.d.). Proteolytic degradation is a known cause of cellulase degradation, as reported by several groups (Dienes, et al. 2007, Hagspiel, Haab and Kubicek 1989).

### 2.1.2 *Pichia pastoris*

*Pichia pastoris* is a methylotrophic yeast from the division of the ascomycota. The subdivision is saccharomycetes with saccharomycetales as class and Saccharomycetaceae as family. *Pichia* is the genus name whereas *P. pastoris* is the species name (Taylor, Spatafora and Berbee 2006).

*P. pastoris* is commonly used as an eukaryotic expression host in research for heterologous protein expression due to different advantages. *P. pastoris* grows on an inexpensive and simple mineral medium. It is capable of generating post-translational modifications such as N- and O-glycosylation and disulfide bonds (Cregg, et al. 2000). It has a strong, inducible promoter and secretes few endogenous proteins, which makes the downstream processing of heterologous secreted proteins much easier (Dosanjh 1996, Cereghino and Cregg 2000, Cregg, et al. 2000).

In most researches where *P. pastoris* or other yeast organisms are used as expression organism, proteolytic degradation has been observed (Cregg, et al. 2000, Sinha, et al. 2005).

## 2.2 Protease

### 2.2.1 Proteases: an introduction

Proteases, which have EC-classification 3.4, hydrolyze the peptide bond. Differentiation of the subclass are documented, there are two sets of subclass peptidases, the exopeptidases (EC 3.4.11-19 and the endopeptidases (EC 3.4.21-24 and EC 3.4.99). The exopeptidases act only near the ends of polypeptide chains, while the endopeptidases act in-between the polypeptide chains. Both sub-subclasses can be differentiated on the basis of catalytic mechanism or substrate specificity. Table 1 gives an overview of different proteases and their classification by mechanism; the substrate specificity is not shown (Moss, E.C. 3.4 acting on peptide bonds n.d.)

Trypsin and chymotrypsin are serine endopeptidases, Trypsin is synthesized as trypsinogen, the zymogen from trypsin, is the inactive protein that is initially transcribed. It is important for a protease to be inactive after synthesis. This way it can only act on the peptide bond where necessary and won't digest the tissues where it's synthesized (McDowall sd).

Table 1: Different proteases with their classification.

Subclasses	Mechanism	Example	EC-classification
carboxypeptidase			
	amino	aminopeptidase I	3.4.11.22
	serine	carboxypeptidase C	3.4.16.5
	cysteine	cathepsin X	3.4.18.1
	metallo	carboxypeptidase A	3.4.17.1
endopeptidase			
	threonine	HsIU-HsIV peptidase	3.4.25.2
	serine	trypsin/ chymotrypsin	3.4.21.1
	cysteine	cathepsin B	3.4.22.1
	metallo	coccolysin	3.4.24.30

Trypsinogen activation is autocatalytic, once a small fraction of trypsinogen is activated, with the help of enteropeptidase to trypsin, trypsin will activate the rest of the trypsinogen (McDowall sd). Trypsin and chymotrypsin are synthesized in the pancreas, where most of the secreting proteins are synthesized (McDowall sd). Chymotrypsin is activated by trypsin-catalyzed cleavage (Voet, Voet and Pratt 2006, Berg, Tymoczko and Stryer 2006).

### 2.2.2 Protease activity in *T. reesei*

Proteolytic degradation of *T. reesei* proteins has been described in several researches. Hagspiel *et al.* detected proteolytic degradation of Cel7A, Cel7B and an  $\alpha$ -galactosidase, which were produced by the modified *T. reesei* strain QM9414 (Hagspiel, Haab and Kubicek 1989).

Eneyskaya *et al.* showed the presence acid protease activity in *T. reesei* medium on  $\alpha$ -galactosidase,  $\beta$ -glycosidase, Cel7A, Cel7B and with a fluorogenic peptide, but the protease itself was not identified (Eneyskaya, *et al.* 1999). Selinheimo *et al.* observed a C-terminal processed tyrosinase in *T. reesei* (Selinheimo, *et al.* 2006). In a more recent study by Stals *et al.* a C-terminal processed endo-N-acetyl- $\beta$ -D-glucosaminidase, secreted by *T. reesei* was also detected (Stals, Samyn, *et al.* 2009).

Contrary, only few studies have been published on the biochemical characterization of *T. reesei* proteases. Dienes *et al.* identified in 2007 a trypsin-like serine protease from this *T. reesei* strain QM9414 (Dienes, *et al.* 2007). The protease activity was shown with zymogram analysis. The protease itself was identified by mass spectrometry after trypsin digestion. Other proteolytic experiments were carried out on BSA and *Hypocrea jecorina* Cel7B. The identified protease showed high homology with fungal serine protease belonging to peptidase family S1 (Dienes, *et al.* 2007).

The trypsin-like protease that Dienes *et al.* identified is different in substrate specificity to the dibasic endopeptidase activity detected by Calmels *et al.* and Goller *et al.* in fungi (Goller,

*et al.* 1998, Calmels, *et al.* 1991). Calmels *et al.* found that several secreted proteins in fungi were processed after a dibasic motif RR, KK, KR and RK (Goller, *et al.* 1998). Several slightly different motifs were also found. Goller *et al.* confirmed this later, they used *T. reesei* Cel7A and Cel7B and two of the fungus xylanases to identify cleavage motifs in *T. reesei*. They identified RA, ER, RR RK and KR amino acid sequence motifs on these proteins (Goller, *et al.* 1998). When we compare this to the data of Dienes *et al.*, we can see that the serine protease cleaves after and arginine or a lysine, but doesn't need a dibasic motif. They found that when arginine or lysine was followed by a proline, the site was almost never recognized as a motif. The substrate specificity of the proteolytic cleaved tyrosinase described by Selinheimo *et al.* and the other detected proteolytic degradation of proteins was not described. The C-terminal processing of an endo-N-acetyl- $\beta$ -D-glucosaminidase secreted by *T. reesei* is in a region and not at one specific place in this protein's amino acid sequence.

### 2.2.3 Protease activity in *P. pastoris*

Proteolytic degradation has been a perpetual problem when yeasts are used as expression organism for recombinant proteins (Cereghino and Cregg 2000). In comparison to other expression hosts, the proteases from *P. pastoris* are not well characterized (Sinha, *et al.* 2005). One of the main reasons for the proteolytic degradation can be found in the oxidative stress and heat shock response due to several external factors, such as the change of carbon source, pH or temperature change. Oxidative stress is in this case a very probable cause, since the methanol metabolism in *P. pastoris* demands high oxygen levels and a form hydrogen peroxide as by-product (Sinha, *et al.* 2005).

Several factors influencing proteolytic degradation in the methylotrophic yeast *P. pastoris* have been studied by Sinha *et al.* with recombinant ovine interferon-T (Sinha, *et al.*

2005). The pH is one of those factors; this can be explained given that each enzyme and protease has a pH-optimum. The use of casamino acids can reduce the proteolysis, perhaps by acting as a preferential substrate for the protease. The effect of the phosphate-level on the proteolysis has been studied by Sinha *et al.* (Sinha, et al. 2005). Results from these studies proved that the protease activity was inversely proportional to the phosphate level. A more important factor for the proteolytic activity is the oxidative stress when *P. pastoris* is grown on methanol. Methanol, which induces the AOX1 promoter for recombinant protein expression, creates oxidative stress conditions (Cereghino and Cregg 2000).

These conditions were tested by comparing the proteolytic degradation when *P. pastoris* was cultivated on methanol and when *P. pastoris* X-33 recombinant cell line was cultivated on glycerol as the sole carbon source. Not only the overall protease activity was measured, but different specific protease activities were measured as well. The protease activities that were measured when methanol was used as substrate were found as well when glycerol was used as substrate, but in much more reduced levels. This proves that the oxidative stress is a very important parameter when recombinant proteins are expressed by *P. pastoris*. A last factor influencing proteolytic degradation is protease inhibition. Different inhibitors belonging to several protease classes were added separately and in combination, to determine their effect on the proteolytic degradation. Sinha *et al.* found that phenyl methyl sulfonyl fluoride (1mM) reduced the total protease activity by 78% and EDTA (1mM) reduced the activity by 45%. When a combination of both (1mM) was used the protease activity was reduced up to 94.2%. In increasing of the concentration of the inhibitors had no decreasing proteolytic effect (Sinha, et al. 2005). The paper by Sinha *et al.* do describe several methods and options to limit proteolytic degradation when *P. pastoris* is used as heterologous expression host.

## 2.3 Endo-N-acetyl- $\beta$ -D-glucosaminidase

### 2.3.1 EC classification

Endo T, expressed by *T. reesei*, is a mannosyl-glycoprotein endo-N-acetyl- $\beta$ -D-glucosaminidase (ENGase) and has EC-classificatie 3.2.1.96. The systematic name for Endo T is: glycopeptide-D-mannosyl –  $N^4$  – (N-acetyl-D-glucosaminyl)<sub>2</sub> – asparagines 1,4 – N – acetyl –  $\beta$  –glucosaminohydrolase (Moss, EC 3.2.1 Glycosidases, i.e. enzymes hydrolysing O- and S-glycosyl compounds 1992, Henrissat, Bairoch and Davies, Glycoside Hydrolase family classification 1998).

### 2.3.2 Glycoside Hydrolase family classification

The CAZy database describes the families of structurally related catalytic and carbohydrate-binding modules of enzymes that degrade, modify, or create glycosidic bonds (Henrissat, Bairoch and Davies, Glycoside Hydrolase family classification 1998). Endo-N-acetyl- $\beta$ -D-glucosaminidase (ENGase) is found in four glycoside hydrolase (GH) families, GH family 18, 20, 85 and 73. In GH family 18 there are two known activities, ENGase (EC 3.2.1.96) and chitinase (EC 3.2.1.14) and several other non-catalytic proteins, such as concanavalin B, narbonin and a xylanase inhibitor (Henrissat, Bairoch and Davies, Glycoside Hydrolase family classification 1998). GH family 85 has only one known activity, which is endo-N-acetyl- $\beta$ -D-glucosaminidase (EC 3.2.1.96) activity and several non-catalytic proteins (Henrissat, Bairoch and Davies, Glycoside Hydrolase family classification 1998). In GH family 20 two 2-domain proteins are present containing both a ENGase GH 18 module and a ... GH 20 module. GH family 73 has several non-classified activities and non-catalytic proteins. Endo-N-acetyl- $\beta$ -D-glucosaminidases (EC 3.2.1.96), a  $\beta$ -1,4-N-acetylmuramoylhydrolase (EC 3.2.1.17) and autolysin (EC 3.5.1.28) are also classified in the GH family 73 (Henrissat, Bairoch and Davies, Glycoside Hydrolase family classification 1998).

Table 2: Endo-N-acetyl- $\beta$ -D-glucosaminidase in the different GH families.

GH family	Eukaryotic / Prokaryotic	Protein name	Organism name	Uniprot code	PDB code
GH 18	Prokaryotic	Endo H	<i>Streptomyces plicatus</i>	P04067.1	1EDT
		Endo F1	<i>Elizabethkingia meningoseptica</i> ATCC 33958	P36911.1	2EBN
		Endo F2	<i>Elizabethkingia meningoseptica</i> ATCC 33958	P36912.1	/
		Endo F3	<i>Elizabethkingia meningoseptica</i> ATCC 33958	P36913.1	1EOK
		BT3987	<i>Bacteroides thetaiotaomicron</i> VPI-5482	O07088	/
		EndoE(*)	<i>Enterococcus faecalis</i> HER1044	Q6U890	/
		EF0114(*)	<i>Enterococcus faecalis</i> V583	/	/
		Endo-Fsp	<i>Flavobacterium</i> sp. SK1022	P80036	/
		EndoS	<i>Streptococcus pyogenes</i> 40/58 PI	Q9PAG4	/
	Eukaryotic	Endo FV	<i>Flammulina velutipes</i>	D1GA49	/
		Endo T	<i>Hypocrea jecorina</i>	C4RA8	(**)
GH 20	Prokaryotic	EndoE(*)	<i>Enterococcus faecalis</i> HER1044	Q6U890	/
		EF0114(*)	<i>Enterococcus faecalis</i> V583	/	/
GH 73	Prokaryotic	Mur2	<i>Enterococcus hirae</i> ATCC 9790	P39046	/
		Acm2	<i>Lactococcus lactis</i> subsp. <i>cremoris</i> MG1363	A2RHZ5	/
		Auto	<i>Listeria monocytogenes</i> EGD-e	Q8Y842	3FI7
GH 85	Prokaryotic	Endo A	<i>Artrobacter protophormiae</i> AKU 0647	Q9ZB22	2VTF
		Endo-BH	<i>Bacillus halodurans</i> C-125	Q9KER4	/
		Endo-D	<i>Streptococcus pneumoniae</i>	Q93HW0	/
		EndoD	<i>Streptococcus pneumoniae</i> TIGR4	Q97S90	2W91
	Eukaryotic	AT3g11040	<i>Arabidopsis thaliana</i>	Q9SRL4	/
		At5g05460	<i>Arabidopsis thaliana</i>	Q9FLA9	/
		Endo-CE	<i>Caenorhabditis elegans</i> Bristol N2	Q19089	/
		Engase1	<i>Homo sapiens</i>	Q8NFI3	/
		Endo-M	<i>Mucor hiemalis</i>	Q9C1S6	/
		Endo-LE	<i>Solanum lycopersicum</i>	/	/

(\*) Both EndoE and EF0114 enzymes are classified in both GH family 18 and GH family 20 according to the CAZy classification of carbohydrate enzymes (Henrissat, Bairoch and Davies, Glycoside Hydrolase family classification 1998) The reason for this dual classification of these enzymes is due to that EndoE and EF0114 have 2 domains, one belonging to GH family 18 and the second to GH family 20 (personal communication by B. Henrissat).

(\*\*) The structure of Endo T has not been published yet.



Table 1 gives an overview of all the different endo-N-acetyl- $\beta$ -D-glucosaminidase found within various glycoside hydrolase families, with their PDB-code, Uniprot code and organism listed. In that table

### 2.3.3 Glycoside hydrolase family 18

The enzymes classified into GH family 18 were initially only coming from prokaryotes. Endo H from *Streptomyces plicatus* (Tarentino, Plummer and Maley 1974), Endo F1, F2, F3 from *Elizabethkingia meningoseptica* (Trimble and Tarentine 1991) and other GH family 18 ENGases are shown in table 2. Eukaryotic ENGases belonging to GH family 18 have only been discovered recently. Endo T, from *Trichoderma reesei* (Stals, Samyn, et al. 2009), and Endo FV, from *Flammulina velutipes* (Hamaguchi, et al. 2010), are the first two eukaryotic ENGase found to be classified into in the GH family 18. All ENGases from GH family 18 are also shown in table 2. Only for four of them, the 3D-structure has been determined. With Endo T as the only fungal ENGase. Typical for them and for all GH family 18 proteins, is their fold. They all have a  $(\beta/\alpha)_8$  TIM barrel structure. This structure is shown in figure 1 and has the form of a cask, where eight  $\beta$ -sheets are placed internally in the barrel anti-parallel with 8 external  $\alpha$ -helices. The 8  $\beta$ -sheets form an inner circle, around which the 8  $\alpha$ -helices form an outer circle. The cavity in the inner circle is apolar due to the apolar side chains of both the  $\alpha$ -helices and  $\beta$ -sheets. The side chains between the  $\alpha$ -helices and  $\beta$ -sheets are dominated by Val, Ile and Leu. The polar amino acids are localized on the surface exposed parts of the barrel, on the top or the bottom, where they contribute to the solubility or the catalytic activity of the enzyme (Branden and Tooze 1999, Voet, Voet. and Pratt 2008). A structural comparison between the characterized GH 18 ENGases is shown in appendix 4.



Figure 1:  $(\beta/\alpha)_8$  TIM barrel structure from Endo T (Boel 2010)

## 2.4 N-Glycosylation

### 2.4.1 N-glycosylation an introduction

Glycosylation, a post-translational modification, is involved in immunogenicity, protein folding and protein stability (Stanley, Schachter and Taniguchi 2009 2nd edition). N-glycosylation is the transfer of a specific sugar chain to a asparagine residue. The transfer of the N-glycan by the ER-oligosaccharyltransferase only occurs in consensus regions consisting of the Asn-x-Ser/Thr motif. X can be any amino acid except a Pro (Karamanos 1997). Different N-glycans are found between organisms and in different cells or tissues, but they can all be classified into three types; high-mannose, complex and hybrid N-glycans (Stanley, Schachter and Taniguchi 2009 2nd edition).

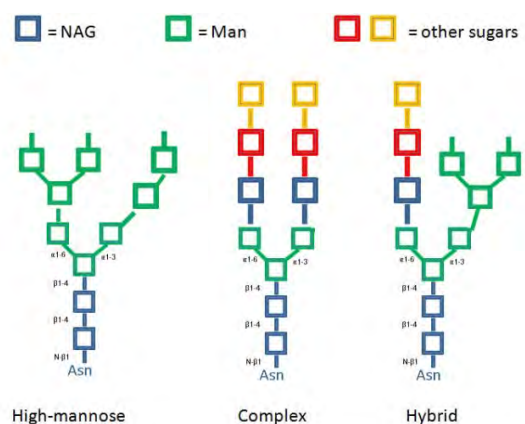


Figure 2: Different N-glycosylation forms (Stanley, Schachter and Taniguchi 2009 2nd edition).

High-mannose glycans consist predominately of mannoses and N-acetyl glucosamines. Complex type n-glycans are built of mannose, N-acetyl glucosamine, galactose, sialic acid and are sub classified according to the number of branching points. Hybrid type N-glycans typically contains two branches; one of the high-mannose type and one of the complex type. The different N-glycans are shown in figure 2.

#### 2.4.2 N-glycosylation in filamentous fungi

Filamentous fungi are often used as expression organism due to their capability of post-translational modifications, such as glycosylation (Deshpande, et al. 2008, Sharma, et al. 2009). In filamentous fungi N-glycosylation is important in secretion and localization of proteins. Only high-mannose glycans are found in filamentous fungi. They can though still vary between different organisms. The degree of mannosylation can vary; from 5-7 mannose residues in e.g. *T. reesei* (Stals, Samyn, et al. 2009) upto 8-9 mannose in e.g. *A. niger* and extra other modifications (e.g. phosphate, sulphate and phosphordiester) can be present (Tsutomu Takayanagia 1994, Deshpande, et al. 2008). Yeast N-glycans can contain more than 100 mannoses on one N-glycan (Lehle 1992).

#### 2.4.3 N-glycosylation in *Trichoderma reesei*

The glycosylation patterns of protein expressed by *T. reesei* can change depending on the cultivation conditions. Other hydrolytic enzymes are expressed and secreted in the cultivation medium and can be active in the correct pH conditions. These exo- and/or endoglycosidases can act on the O- and N-glycans present on the glycoproteins (predominantly cellulases). These post-secretorial modifications of both O- and N-glycosylations have been studied extensively on Cel7A in *T. reesei* (Stals, Sandra and Geysens, et al. 2004, Stals, Sandra and Devreese, et al. 2004). And the activity of a  $\alpha$ -

mannosidase, phosphatase and endo-N-acetyl- $\beta$ -D-glucosaminidase activity was observed (Stals, Samyn, et al. 2009). Two genes for endo-N-acetyl- $\beta$ -D-glucosaminidases, one intracellular and a second extracellular ENGase (Endo T) are present in the *T. reesei* genome (Martinez D 2008). Endo t was in a later study purified from the culture broth and partially characterized (Stals, Samyn, et al. 2009).

### 2.5 Endo T

#### 2.5.1 Substrate

*T. reesei* Endo T hydrolyzes the  $\beta$ -1,4 glycosidic bond between the two N-acetyl glucosamine (NAG) of high-mannose N-glycans as described above. Therefore one NAG stays bound to the protein after the removal of the N-glycan by Endo t. The second NAG remains bound to the removed high-mannose N-glycan (Karamanos 1997). The presence of the enzyme in the medium explains for the occurrence of single NAG residues on *T. reesei* proteins as reported by several groups (Stals, Sandra and Geysens, et al. 2004).

#### 2.5.2 Mechanism

Endo T and all ENGases are retaining glycoside hydrolases using neighboring group participation for the catalytic mechanism. The distance between the catalytic residues Asp 129 and Glu 131 in the retaining mechanism is approximately 10 Å. The mechanism is shown in figure 3. There has to be 1 acetic residue being an Asp or Glu and one stabilizing group. For all ENGases in GH family 18 the catalytic residues are Asp and Glu with only one residue between them. The acetic residue interacts with the substrate and forms an oxazoline intermediate, this is formed after a transition state. A second acetic group stabilizes the intermediate and brings the active site back to its original state (Davies and Henrissat 1995).

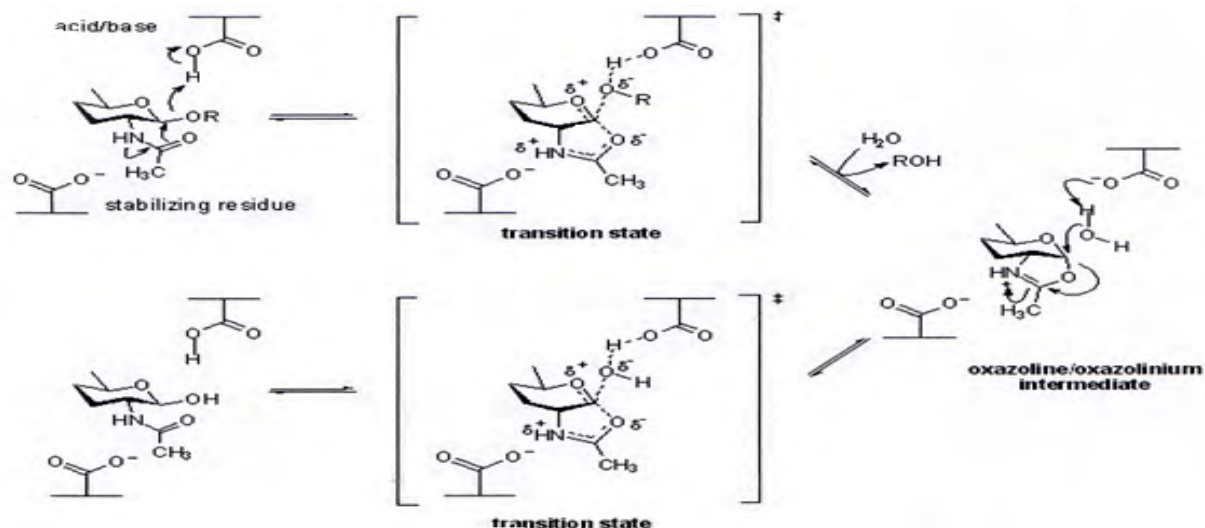


Figure 3: Retaining mechanism with a N-acetyl or N-glycol group on position 2 (Davies and Henrissat 1995).

### 2.5.3 Structure

The structure of the active form of the endo-N-acetyl- $\beta$ -D-glucosaminidase secreted by *T. reesei*, i.e. Endo T, has been solved by X-ray crystallographic methods and the structure was refined to 1.3 Å. The protein was crystallized using a protein concentration of 8.3 mg/ml, in 0.1 M citric acid buffer, with a pH between 2.4-3.0 and using PEG 3350 or PEG1500/PEG8000 as a precipitant, with both PEG concentrations varying from 4 to 14 % and by incubating the experiments at 20° C (Digre 2010). The tertiary structure of Endo T is an  $(\beta/\alpha)_8$  TIM barrel and the structure has overall dimensions of approximately 45 Å x 34 Å x 57 Å. A cartoon representation of the Endo T structure is shown in figure 4.



Figure 4: A cartoon representation of the Endo T structure with a ligand bound in the active site of the enzyme, on top of the B-barrel. The bound ligand is colored yellow, the catalytic residues are colored blue and the two bound NAG residues red

## 2.6 X-ray crystallography

### 2.6.1 Protein crystallography

Protein crystallography has been thought to be more an art than a science for a very long time (Chayen, Turning protein crystalization from an art into a science 2004). A lot of patience, perseverance, intuition and luck is needed to grow high-quality protein crystals (Navarro, Wu and Wang 2009). Nowadays more and more proteins are routinely crystallized. This is mainly due to the high-throughput crystallization methods that have been developed over the past decade (Chayen, Protein crystallization strategies for structural genomics 2007).

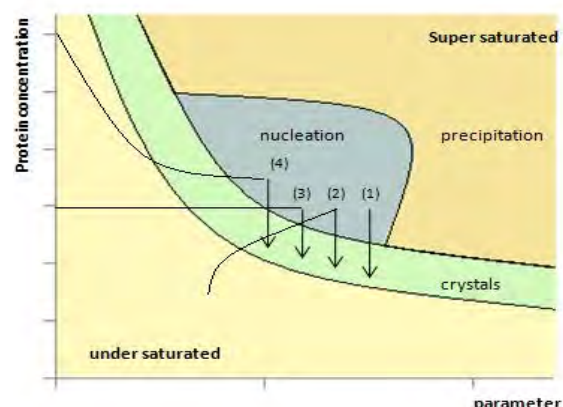


Figure 5: Protein crystallization phase diagram. The different paths are show for the different crystallization methods; (1) batch method, (2) vapor diffusion, (3) dialysis and (4) free interface diffusion (Chayen, Turning protein crystalization from an art into a science 2004).



Proteins molecules in supersaturated state can precipitate, nucleate or form crystals. Figure 5 shows the phase diagram for all these different states of a protein. Only after nucleation, protein crystals start to be formed and grow. Crystal growth occurs in the meta-stable zone, shown on Figure 5. It is the level of supersaturation, which is the main drive for a protein to crystallize. For X-ray crystallography big high-quality crystals are needed to obtain a good x-ray diffraction pattern when the protein crystals are exposed to x-ray beam. Massive nucleation and precipitation should thus be limited (Chayen, Turning protein crystalization from an art into a science 2004). Different techniques can help to overcome this problem when normal methods fail to crystallize proteins (Chayen, Methods for seperating nucleation and growth in protein crystalization 2005). Examples of these methods are e.g. adding nucleation seeds (both macro and micro seeding) to a crystallization drop and dilution of the protein solution used for the crystallization experiment. The difference between macro and micro seeding is that in the former one introduce single crystals to the crystallization drop while in the later case one try to ad controlled amounts of pre-grown nucleation sites to the crystallization drop. In both cases with the hope that these seeding strategies will give raise to bigger and better quality crystals of the target protein (Chayen, Methods for seperating nucleation and growth in protein crystalization 2005).

Different parameters influence a crystallization experiment, these can be categorized in ether bio-physical phenomena or crystallization methods (Bergfors 1999). Bio-physical parameters are for instance protein concentration, temperature, pH, the precipitant and its concentration, the salt and its concentration and so on. Different thermodynamic and kinetic parameters for a crystallization experiment could in theory be calculated, but would create a lot of work and these parameters will be different for each new protein (Navarro, Wu and Wang 2009).

The importance of the different crystallization methods should not be underestimated. A different path to get to the meta-stable zone and to grow high-quality crystals characterizes each method. These different paths are shown in Figure 5 and all of these methods will pass the nucleation phase, before dropping to the meta-stable zone and crystal growth (Chayen, Turning protein crystalization from an art into a science 2004). The batch method was the first method that was developed for protein crystallization. It is an easy and simple method, still in use but not so common any longer (Bergfors 1999). Various types of so called vapor diffusion methods are now more commonly used. In the vapor diffusion crystallization method different setups are available. These different setups are hanging drop, sitting drop and sandwich drop and are shown in Figure 6. The hanging drop vapor diffusion method is commonly used to set up manually crystallization experiments. The sitting drop method is most often used to setup crystallization plates with various types of crystallization robots. For the hanging drop method a volume of 500 µl is usually used in the crystallization experiment reservoir containing the crystallization solution. This crystallization solution contains buffer (concentration often varying between 0.1 and 0.2 M), metal ions and crystallization precipitant, all in different concentrations depending on the crystallization condition. One of the most commonly used precipitants in a protein crystallization experiment are different types of poly ethylene glycol (PEG). Their average molecular weight can vary.

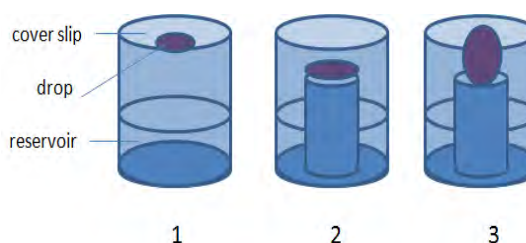


Figure 6: Different vapour diffusion methods. Hanging drop (1), sitting drop (2) and sandwich drop (3) method (Bergfors 1999).

There are two steps to get high-quality crystals. Screening to identify initial crystallization conditions is the first step. Different companies sell crystallization screens that are used to identify the initial crystallization condition for the protein of interest. An example of a crystallization robot is the Oryx robot from Douglas instruments, Hungerford (Bergfors, Terese M. 2009). Two examples of commercially available crystallization screens are; the JCSG or the Core 96 screen and the PEG ION screen (HAMPTON 2003, QIAGEN 2003). Crystallization robots are very useful tool to save time and reproduces crystals in a higher ratio than manual crystallization setups do.

When initial crystal conditions are found, these crystals have to be tested to determine whether these crystals are protein or salt crystals. If the initial crystals are found to be protein crystals the crystallization condition is further refined to produce bigger crystals. This is the second step in a crystallization process and is often called optimization of the crystallization condition. Different ways of identifying initial crystallization conditions and improving this further to produce high quality protein crystals are well described in the book: "Protein crystallization second edition, edited by Terese M. Bergfors". Other techniques that have not been mentioned, can also be found in this handbook (Bergfors, Terese M. 2009). When protein crystals are big enough to be used for an X-ray data collection, these are picked out the crystallization drop. The crystal is moved to a cryo solution and frozen in liquid nitrogen. The function of keeping the crystal frozen at a cryogenic condition is to prevent that the crystal is thermally destroyed in the strong X-ray beams in use today, often strong synchrotron X-ray sources To identify a good cryo-solution can be very tidies and difficult. Therefore several conditions should be tried (Bergfors, Terese M. 2009). After freezing the crystals, these can be stored in liquid nitrogen before being transported to the synchrotron, where the crystals will be exposed in the X-ray beam.

### 2.6.2 X-ray diffraction

The longest and most difficult part to get a good 3D model, is getting high-quality crystals. After subjecting the crystal to X-ray's, the diffraction pattern has to be translated in to a 3D structure model. Depending on the size of the protein and the quality of the diffraction pattern obtained when exposing the crystal in the X-ray beam, solving the structure and refining the structure model it is the second biggest task. At a synchrotron the protein crystals are exposed in an X-ray beam with a wavelength of approximately 1 Å. A X-ray data collection can consist of anything between hundred up to many hundreds images, where the oscillation of the crystal in the x-ray beam and the exposure time can be adapted to achieve as god diffraction as possible from the exposed crystal.

### 2.6.3 Electron density map

The first task after a X-ray data collection is to translate the diffraction pattern into an electron density map. Crystals, which are subjected to X-rays, will give a specific diffraction pattern. This diffraction is the result of enhanced X-ray radiation in some directions and extinguished X-ray radiation in other directions. Each electron of the atoms building up the molecules of the crystal will influence the final diffraction pattern obtained, since each reflection will change the intensities of other reflections. A crystal is a symmetric repetition of the molecule building up the crystal, (Lattman and Loll 2008).

Each reflection is characterized by its position, amplitude and its phase. Both the amplitude of all diffraction spot of a X-ray diffraction dataset and the phases information for each of these diffraction spots are necessary to know to be able solve the three dimensional structure of a protein using the diffraction dataset obtained. The amplitudes can directly be derived from the intensity of the reflection recorded on the diffraction image collected when exposing the crystal in the x-ray beam. The phases on the other hand cannot be derived directly.

Several techniques are available to determine the phases for the recorded reflections. Determination of the phases by isomorphous replacement, and phasing with a highly similar known structures (molecular replacement) are two techniques which are commonly used to derive the phase information for the collected reflections (Lattman and Loll 2008). Once the phase information for the diffraction spots are known the diffraction pattern will be translated into an electron density maps. This translation is done by computer software

#### 2.6.4 Protein structures

A structure model of the protein that is studied is modeled into the electron density map calculated from the diffraction pattern obtained. Refining a protein model is the modeling in of the individual amino acids of a protein into the calculated electron density map. Usual during refinement of a structure two different electron density maps are calculated from a diffraction pattern, the (2Fo –Fc) and the (Fo-Fc) map.

The quality of the built structure model depends a lot on the on the quality of the calculated electron density map and by the refinement and fitting of the model in the electron density map. The higher resolution of the diffraction pattern one can collect, the better the electron density map gets. An electron density map with a resolution of 3 Å will show the main structure elements of the protein, an electron density map of approximately 1 Å resolution will show the protein studied at atomic resolution, including the hydrogen atoms of the protein. The higher the resolution of a collected diffraction pattern gets, the more reflections the collected dataset contains (Wlodawer, et al. 2007). A second parameter, which is a control for the quality of the protein model are the R and R<sub>free</sub> factors. These factors measure the difference between the observed and calculated amplitudes of the electron density. The lower these R-factors are the better the quality is. There is though one general remark, the value of the R-factors cannot keep dropping, they are correlated to

the quality of the electron density map. The lower the difference between the R and R<sub>free</sub> the less difference there is between the observed and calculated amplitude and the better the model is. The angles of the phases are also a parameter, which should be checked regularly (Wlodawer, et al. 2007).

The amino acids of a protein are modeled into the electron density map calculated from the collected diffraction pattern. Not all geometric conformations of the peptide bond between two connected amino acid in the poly-peptide chain of a protein are allowed. Depending on the quality of the electron density map, the restraints for the geometric parameters for the individual amino acids building up a protein and the geometric for the angles between two connected amino acids in the poly-peptide chain can be loosened up gradually, and the dependents between the structure model and the electron density map can be increased. The refinement of a structure model can be very time consuming and will depend a lot on the resolution and the quality of the calculated electron map for the protein. The better the phases get, the better the structure model will be. The phases for the calculated electron map can be improved by adding solvent molecules to the structure model in addition to adding the amino acids for the protein part of the model. Several programs are capable of refining structures and they are the perfect tools to solve protein structures.

### 3 Materials and methods

#### 3.1 *Protein samples*

The intact Endo T samples were heterologously expressed in *P. pastoris*. A mutated Endo T (D129A) protein expressed in *T. Reesei* was also used, but we weren't able to confirm if it was catalytic dead.

#### 3.2 *Endo T protein analysis*

##### 3.2.1 *SDS-PAGE analysis*

Samples of expressed and purified Endo T proteins were checked using 4-20% (15 well) gradient gels or on 12% (15 well) homogenous poly-acrylamide gels from Bio-Rad. The gels were bought pre-casted from the Bio-Rad. 5  $\mu$ l of the precision standard ladder was loaded on gel, while 12  $\mu$ l of the treated protein sample was loaded. Protein samples were treated by adding 5  $\mu$ l of SDS dye sample buffer to 15  $\mu$ l of non-treated protein samples and heated for 5 minutes at 95°C. The gel was run for 40 minutes with a SDS containing running buffer and washed several times with water before being stained. After washing, the gels were stained for one hour with Bio-safe G-250 stain from Bio-Rad and the gels were destained overnight in water.

##### 3.2.2 *Native-PAGE analysis*

Endo T proteins were treated as the SDS-PAGE analysis, with the only difference that the protein sample was loaded untreated. The gel was run for 3 hours minutes with native running buffer and washed several times with water before being stained. After washing the gels were stained for one hour with Bio-safe G-250 stain.

##### 3.2.3 *Peptide fingerprinting by mass spectrometry*

Several bands were analyzed with mass spectrometry after trypsin digestion, to identify if the detected band on the gel corresponded to the intact or processed form of Endo T. All the mass spectrometry experiments were carried out using a Ultraflex MALDI TOF/TOF, Bruker Daltonics and carried out by Åke

Engström at the Department of medical biochemistry and microbiology, biomedical centre Uppsala, Sweden.

##### 3.2.4 *Western Blot*

SDS-PAGE gels were run in duplicate, one was stained as described above, while the proteins of the unstained gel were transferred on a nitrocellulose membrane as described by Burnette (Burnette 1981). After the transfer, the bands containing terminal mannoses can be detected with the glycan differentiation kit from Boehringer Mannheim Biochemica as described in the DIG kit procedure (Biochemica 1996). But the transfer of the proteins to the nitrocellulose filter failed twice. The gels are shown in appendix 3.

##### 3.2.5 *Protein concentration*

The absorbance of the Endo T protein sample was first measured using a Nanodrop spectrophotometer. Samples of 2  $\mu$ l protein were used to determine the protein concentration using the Nanodrop 2000. The protein concentration was measured at 280 nm, and the absorbance at 260 nm was also measured to determine the concentration of DNA in the sample. The absorbance was then used to calculate the protein concentration with the specific protein extinction coefficient.

When a bad 260/280 value was measured, the absorbance couldn't be used to determine the total protein concentration of the sample. The total protein concentration was then measured by using the Bio-Rad protein assay, which is based on the Bradford method (Bio-rad n.d.). 20  $\mu$ l of a bovine serum albumin standard (from 0 to 0.8 mg/ml) was incubated with 1 ml of 5 times diluted dye reagent (diluted in de-ionized water) at room temperature for 5 minutes. The same procedure was carried out applied for the dilution series of the Endo T protein sample. The concentration of the Endo T protein samples were then calculated back from the absorbance. All measurements were conducted in triplicates and measured at  $\lambda=595$  nm.

### **3.3 Purification of the intact Endo T for crystallization**

#### **3.3.1 Size-exclusion purification with a 3 kDa cutoff filter.**

The extracellular culture medium from a 500 ml *Pichia pastoris* culture was first Buchner filtered to get rid of cell debris. Then the culture medium was concentrated with a 3 kDa cutoff filter to a final volume of approximately 12 ml in 20 mM pH 8 Tris buffer, protease inhibitors were added to keep Endo T intact. This Endo T protein sample was checked on SDS-PAGE. Patricia Ntarima conducted the first purification steps of Endo T at the university college Ghent.

#### **3.3.2 Superdex 75 size-exclusion chromatography**

The intact Endo T with protease inhibitors added to the protein solution was purified by size-exclusion chromatography. The column was first cleaned with water and thereafter pre-incubated with 100 mM sodium chloride 20 mM pH 5 sodium acetate buffer. 500 µl of 9.40 mg/ml Endo T sample was loaded to the column and eluted with 100 mM sodium chloride 20 mM pH 5 sodium acetate buffer after 15 minutes. The concentration of the Endo T was measured using the protein assay from Bio-Rad. The fractions from the peaks were checked on a 4-20% gradient gel SDS-PAGE after first being concentrated to approximately 750 µl with a 2 ml Vivaspin column with a molecular cutoff filter of 3kDa.

#### **3.3.3 MonoQ anion exchange chromatography**

The monoQ column was first washed with water, with 2ml 2M NaCl, 2ml 1 M NaOH, 2ml 1M HCl and 2 ml of 75 % acetic acid to get rid of all impurities before using it. The column was cleaned with water and with a strong buffer before calibrating the column. The calibration of the column was done by first running 40 ml 20 mM pH 6.5 Bis-Tris buffer and with approximately 40 ml 1 M NaCl 20 mM pH 6.5 Bis-Tris buffer. 20 ml 20 mM pH 6.5

Bis-Tris buffer was brought over the column before loading the protein sample on the column. 1 ml of Endo T in 20 mM pH 8 Tris buffer was diluted in 49 ml of 20 mM of Bis-Tris buffer to change the pH to 6.5 and to lower the conductivity of the sample. The protein was eluted with a linear gradient from 0 % to 100% (1 M NaCl 20 mM pH 6.5 Bis-Tris buffer). Changing the gradient from 0 % to 70 % gave a better separation of the different eluted peaks. The peak fractions were concentrated with Vivaspin columns with a molecular weight cut-off filter of 3kDa and then checked on 12 % homogenous SDS-PAGE. The protein concentration was determined using the Nanodrop.

#### **3.3.4 Deglycosylation of intact Endo T for crystallization**

A sample of intact Endo T, in 20 mM pH 8 Tris buffer with protease inhibitors, was buffer exchanged against 50 mM pH 5 sodium acetate buffer. The sample was then incubated at 37°C for approximately couple of days to let the active fraction of Endo T present in the sample deglycosylated all N-glycans bound to the inactive form of Endo T. The results of the deglycosylation experiments were checked on 4-20% gradient SDS-PAGE. The deglycosylated fractions could then be used for crystallization experiments.

### **3.4 Proteolytic stability of the intact Endo T**

The protease inhibitor cocktail used is the Halt™ Protease Inhibitor Single-Use Cocktail EDTA-Free from Thermo Scientific. 10µl of the cocktail was added to 1ml of the protein sample as described in the protocol for the used protease inhibitors.

#### **3.4.1 pH-stability**

150 µl of Endo T samples with a concentration of 9.40 mg/ml, both with and without protease inhibitors, were incubated at 21°C in 100 mM citric acid buffer pH 3, 100 mM acetic acid buffer pH 5, 100 mM TRIS pH 8 and 100 mM pH 11.61 citric acid phosphate buffer to check their proteolytic pH stability. Six different



samples of all pH conditions were taken in a period between 5 hours and 14 days after incubation and checked on 12% homogenous SDS-PAGE.

### *3.4.2 Temperature stability*

150 µl of 100 mM pH 5 sodium acetate buffer Endo T samples with a concentration of 9.40 mg/ml, both with and without protease inhibitors, were incubated at 4, 21 and 37 °C to check their proteolytic temperature stability. Six different samples of all conditions were taken in a period of 14 days and checked on 12% homogenous SDS-PAGE.

### *3.4.3 Proteolytic degradation*

Both the wild type Endo T and the mutant Endo T (D129A) were incubated for 1 day at 20°C in a 0.1M pH 3 citric acid buffer. Protein samples were taken at time 0, 2 hours, 6 hours, and 1 day. A mixture of 1/1 was also incubated in the same conditions. The different protein samples were then checked on a 12 % homogenous SDS-P.

### *3.4.1 Proteolytic activity*

Both the wild type Endo T and the mutant Endo T (D129A) were incubated for 1 day at 20°C in 0.1M pH3 citric acid buffer, and thereby lose the C-terminus of the protein. These samples were then incubated with T. reesei Cel7A for 2 hours at 20°C in 0.1M pH3 citric acid buffer. The incubation was performed at different Endo T/Cel7A weight ratios of 1/1, 1/10, 1/20, 1/40 and 1/100. The non-cleaved and cleaved form of Cel7A, by papaine digestion were taken as control samples. A 1/1 ratio of both the wild type Endo T and the mutant Endo T (D129A) were incubated with Cel7A for 2 hours at 20°C in 0.1M pH3 citric acid buffer, without losing their C-terminus.

## *3.5 X-ray crystallography*

### *3.5.1 Protein crystallization*

Crystallizations were setup up at 20°C, both for the hanging drop and sitting drop plates. The Oryx robot, from Douglas instruments, was

used to set up crystallizations with the JSCG and the PEG ION crystallization screens (QIAGEN 2003, HAMPTON 2003). The well volume of the crystallization experiments was 70 µl . The total drop volume of the crystallization experiment was 0.70 µl. The drop consisted of 0.35 µl well solution, and of 0.35 µl protein sample. When hanging drop crystallization plates were setup the well volume contained 500 µl and the crystallization drop had a final volume of 4µl. The drop consisted of 2µl well solution and 2 µl protein sample. The well liquid consisted of a 0.100 M buffer, the pH and buffer was varied and PEG 3350, where the concentration was varied. In both setups, the protein concentration was varied.

### *3.5.2 X-ray diffraction*

The Endo T crystals were transported to the synchrotron (MAX-Lab, Lund or ESRF, Grenoble) where they were tested in the X-ray beam at the synchrotron . The oscillation range for the data collection was 1°, and the number of images collected depended on the strategy used for the data collection.

### *3.5.3 Protein structure and refinement*

The different protein structures were solved by processing all images using the program imosflm (Powell 2010). When the collected diffraction images were processed, the space group of the diffraction pattern had to be determined. When this was done. The processed diffraction data was then scaled with the scala program within the ccp4 packages (International Union of Crystallography 2011). After scaling the obtained dataset an already available Endo T model was refined against this new dataset. Waters and other components were built in to the structure model to further improve the phases of the electron density map. The R and Rfree values, the bond distance and bond angles were checked after each refinement run to obtain a good structure model.

## 4 Results

### 4.1 *Endo T* protein analysis

#### 4.1.1 *Endo T* transport control

The protein samples with expressed *Endo T* were transported from the University College Ghent to SLU, Uppsala; there they were analyzed using a 4-20% (15 well) gradient SDS-PAGE gel to check if the *Endo T* protein in these were still intact. Both the sample with inhibitors and the sample without inhibitors did still contain intact *Endo T* protein after transport. If lower bands appear proteolytic degradation or deglycosylation takes place. The short processed *Endo T* was also loaded on gel as a reference. These results are not shown.

Lane 2 to 4 :Crude processed *Endo T* sample

Lane 5 to 7: Crude intact *Endo T* sample – inhibitor

Lane 8 to 10: Crude intact *Endo T* sample+ inhibitor

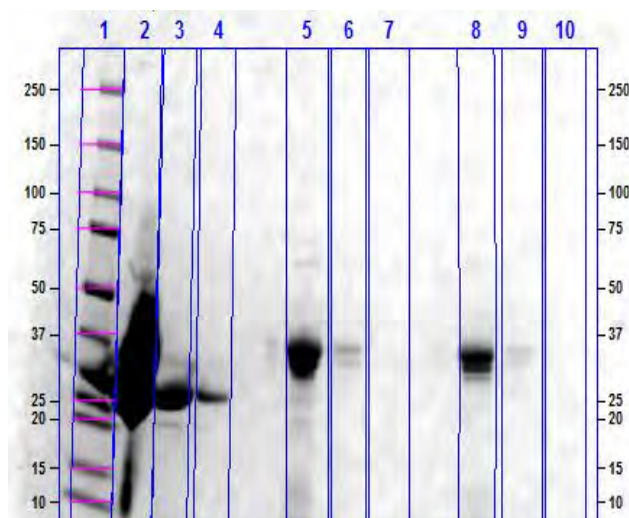


Figure 8: Protein status of *Endo T* (undiluted; 10 times and 100 times diluted) was checked after transport. Lane 2 to 4 contain the dilution series of the active *Endo T*, lane 5 to 7 the *Endo T* sample without inhibitors and lane 8 to 10 contain the *Endo T* sample with inhibitors.

#### 4.1.2 Protein concentration

The total concentration of the crude *Endo T* sample with protease inhibitors was approximately 9.40 mg/ml, the sample without protease inhibitors had a concentration of 2.84 mg/ml. These results are shown in appendix 1.

## 4.2 Purification of the intact *Endo T* for crystallization

### 4.2.1 Size exclusion chromatography

Three batches of 0.500 ml *Endo T* with protease inhibitors were purified with a superdex75 column. The purification by size exclusion failed. These results are shown in Figure 7. No further purification with this method was used. In the first trial no salt was added and the protein was lost, in the second trial the protein purification failed and in the third trial *Endo T* got degraded.

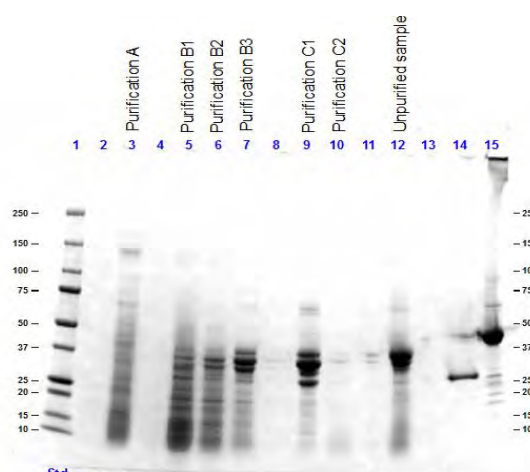


Figure 7: Size exclusion chromatography of *Endo T* with protease inhibitors. The first purification (A) without any salt in the elution buffer failed and the protein was lost. The two other attempts (B and C) with salt, failed or the protein got degraded.

### 4.2.2 Anion exchange chromatography

A batch of 1 ml *Endo T* with protease inhibitors was purified with anion exchange chromatography using a MonoQ column. Figure 10 shows the chromatogram from the separation with a linear gradient of 1 M NaCl from 0 to 70%. Different fractions corresponding with the different peaks were then checked by SDS-PAGE, to identify the *Endo T* fractions containing *Endo T*. These results are shown in Figure 11. The separation was very good and was even better with a gradient from 0 to 70%. The concentrated protein samples are shown in appendix 2.

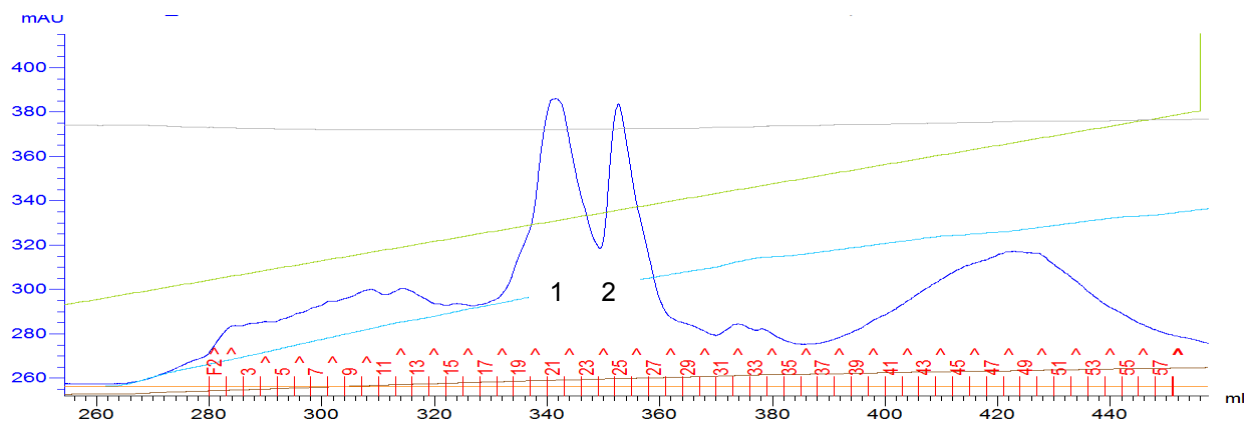


Figure 10: Chromatogram of anion exchange chromatography with a MonoQ column and a 1 M NaCl gradient going from 0 to 70 %. The gradient consists of 20 mM Bis-Tris pH 6.5 and 1M NaCl 20 mM Bis-Tris pH 6.5. Peak 1 around 340 ml is the first fraction; peak 2 around 355 ml is the second fraction containing a different form of the intact Endo T.

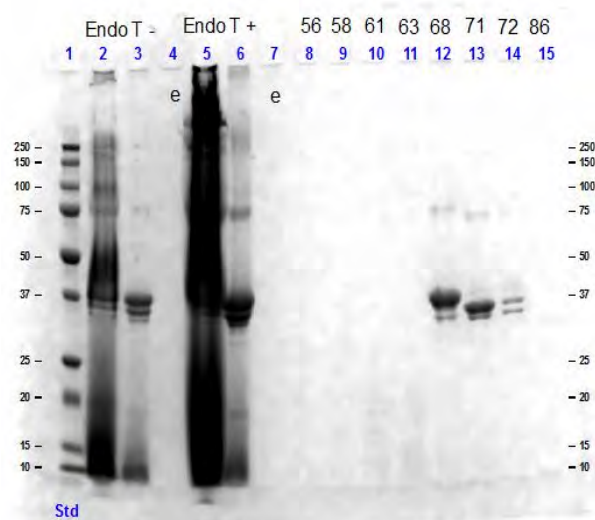


Figure 11: Different fractions of the first MonoQ purification. Lane 2, 3 and 5, 6 contain unpurified Endo T as control. Samples 56 to 63 correspond with the first small peak. Fraction 68 with peak 1 around 340 ml in figure 3 and fractions 71 and 72 with peak 2 around 355 ml in figure 3. Fraction 86 is from the last peak. These fractions were concentrated after purification.

#### 4.2.3 Deglycosylation of intact *Endo T* for crystallization

Deglycosylation of the intact Endo T, by the active fraction of Endo T, in the unpurified protein sample, is shown in Figure 9. During deglycosylation at 37 °C a pull down of protein was noticed. The sample was centrifuged and the supernatant was removed. The pull down was resuspended in a 50 mM pH 5 sodium acetate buffer and also checked on SDS-PAGE gel. The pull down was identified as Endo T. This is though the first time that a pull down of Endo T is noticed.

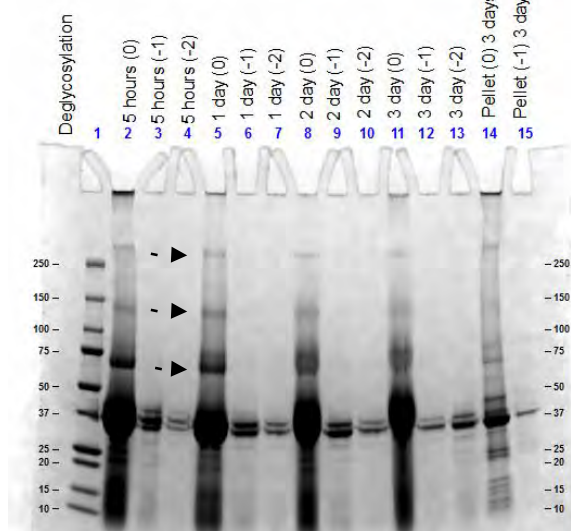


Figure 9: Deglycosylation of inactive Endo T by the active Endo T fraction from the unpurified sample. The values between brackets represent the log scale of the dilution. Higher bands around 75 kDa, 150 kDa and higher than 250 kDa were identified as Endo T.

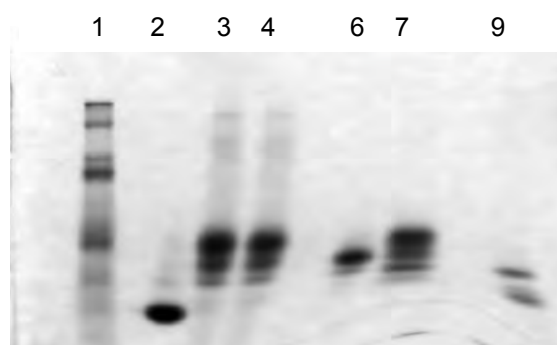


Figure 12: The resulting gel after Native polyacrylamide gel run to check if oligomers of Endo T are formed. Lane 2 contains the active Endo T, lane 3 and 4 the intact Endo T, both unpurified (3) and purified (4). Lane 6 and 7 contain anion exchange fractions (F1Q1 and F2Q2) of the intact Endo T, lane 9 contains the deglycosylated protein sample shown in figure 9.



The deglycosylated sample was incubated with protease inhibitors at 37°C in 0.1 M pH 5 sodium acetate buffer for over a week and was kept intact. After loading the different deglycosylation samples on gel, new prominent bands were detected. These bands around 75 kDa, 150 kDa and higher than 250 kDa were cut out and were subjected to mass spectrometry for peptide identification. The bands were identified as Endo T and could've been oligomers or highly glycosylated forms of Endo T. A native gel was run and no oligomers were visible. The concentration used in the Native PAGE is lower than the detected Endo T bands around 75 kDa, 150 kDa and higher than 250 kDa. Because of this we cannot conclude whether these higher bands are oligomerizations or not.

### 4.3 Proteolytic stability of the intact Endo T

#### 4.3.1 T- dependent degradation pattern of Endo T

Proteolytic degradation of intact Endo T by protease activity was observed in protein samples incubated at different temperatures. In lane 13 of figure 13 there are 5 bands, proteolytic degradation is confirmed when band 4 and/or band 5 is observed. At 4°C proteolytic degradation was observed after 8 days. At 21 °C degradation was observed after 6 days and at 37 °C already after 3 days. Figure 13 to 15 show the degradation pattern at the three different temperatures. The negative control, the Endo T sample with protease inhibitors didn't work.

Degradation was detected in samples of Endo T with and without protease inhibitors present. The reason for this is that the protein sample with protease inhibitors was kept in the freezer and it was not until a later stage the we realized that the protease inhibitors lose their effect when kept frozen at -20°C for a longer period of time. No new inhibitors were added before setting up the experiment. For this reason the negative control failed. In previous experiments the protease inhibitors had been tested and had worked.

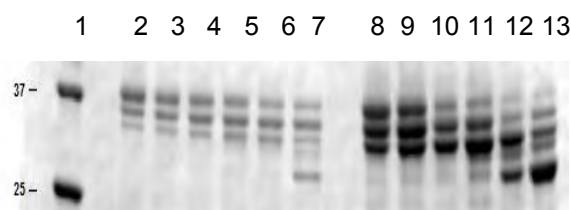


Figure 13: SDS-PAGE showing the degradation pattern of Endo T incubated at 4°C in 100 mM pH 5 sodium acetate buffer. Samples were taken after 5 hours; 1 day; 3 days; 6 days; 8 days and 14 days. Lane 1 is the molecular ladder, lane 2 to 7 contain the Endo T protein sample at the different time points without the protease inhibitor cocktail, lane 8 to 13 contains the Endo T protein sample at the different time points with the protease inhibitor cocktail.

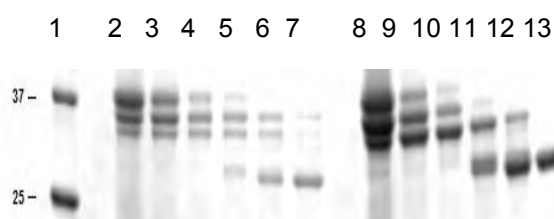


Figure 14: SDS-PAGE showing the degradation pattern of Endo T incubated at 21°C in 100 mM pH 5 sodium acetate buffer. Samples were taken after 5 hours; 1 day; 3 days; 6 days; 8 days and 14 days. Lane 1 is the molecular ladder, lane 2 to 7 contain the Endo T protein sample at the different time points without the protease inhibitor cocktail, lane 8 to 13 contains the Endo T protein sample at the different time points with the protease inhibitor cocktail.

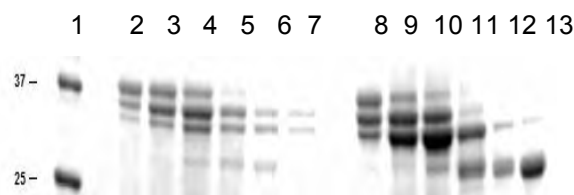


Figure 15: SDS-PAGE showing the degradation pattern of Endo T incubated at 37°C in 100 mM pH 5 sodium acetate buffer. Samples were taken after 5 hours; 1 day; 3 days; 6 days; 8 days and 14 days. Lane 1 is the molecular ladder, lane 2 to 7 contain the Endo T protein sample at the different time points without the protease inhibitor cocktail, lane 8 to 13 contains the Endo T protein sample at the different time points with the protease inhibitor cocktail.

### 4.3.2 pH-dependent degradation pattern of Endo T

Proteolytic degradation by protease activity was observed in all different pH conditions except the one at pH 11.63. The samples at this pH are not so visible compared to protein samples in other conditions. For that reason it is not possible to draw a direct conclusion about proteolytic cleavage of Endo T at this high pH condition. There was both degradation in the sample with and without protease inhibitors. The reason therefore is described above under Temperature-dependent degradation pattern of Endo T.

Protein degradation is accelerated when pH is decreased. Figure 16 to 19, show the degradation pattern of Endo T. At pH 3 full degradation of the protein is observed after one day. A second experiment was setup to give a better image of the degradation of Endo T at pH 3. In this experiment fresh protease inhibitors were added to the protein sample before incubation. At pH 5 degradation of Endo T occurred after 6 days. The same period is needed at pH 8, but the amount of degraded Endo T is lower than at pH 5. No direct conclusion can be drawn about the degradation of Endo T incubated at pH 11.6.

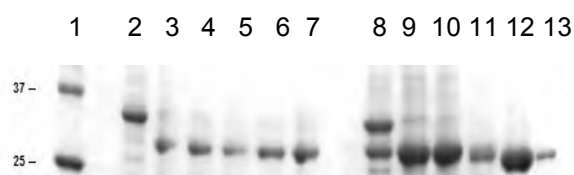


Figure 16: SDS-PAGE showing the degradation pattern of Endo T incubated at 21 °C in 100 mM citric acid buffer pH 3. Samples were taken after 5 hours; 1 day; 3 days; 6 days; 8 days and 14 days. Lane 1 is the molecular ladder, lane 2 to 7 contain the Endo T protein sample at the different time points without the protease inhibitor cocktail, lane 8 to 13 contains the Endo T protein sample at the different time points with the protease inhibitor cocktail



Figure 17: SDS-PAGE showing the degradation pattern of Endo T incubated at 21 °C in 100 mM acetic acid buffer pH 5. Samples were taken after 5 hours; 1 day; 3 days; 6 days; 8 days and 14 days. Lane 1 is the molecular ladder, lane 2 to 7 contain the Endo T protein sample at the different time points without the protease inhibitor cocktail, lane 8 to 13 contains the Endo T protein sample at the different time points with the protease inhibitor cocktail.

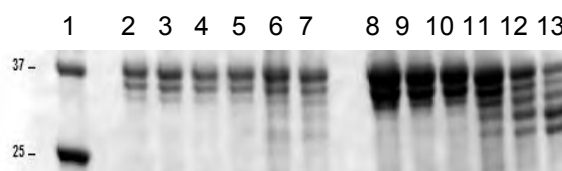


Figure 18: SDS-PAGE showing the degradation pattern of Endo T incubated at 21 °C in 100 mM Tris buffer pH 8. Samples were taken after 5 hours; 1 day; 3 days; 6 days; 8 days and 14 days. Lane 1 is the molecular ladder, lane 2 to 7 contain the Endo T protein sample at the different time points without the protease inhibitor cocktail, lane 8 to 13 contains the Endo T protein sample at the different time points with the protease inhibitor cocktail.

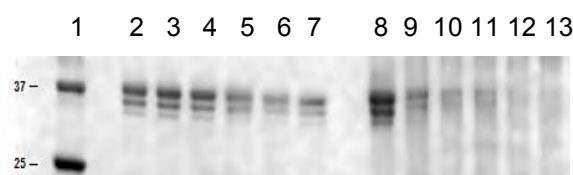


Figure 19: SDS-PAGE showing the degradation pattern of Endo T incubated at 21 °C in 100 mM phosphate-citric acid buffer pH 11.6. Samples were taken after 5 hours; 1 day; 3 days; 6 days; 8 days and 14 days. Lane 1 is the molecular ladder, lane 2 to 7 contain the Endo T protein sample at the different time points without the protease inhibitor cocktail, lane 8 to 13 contains the Endo T protein sample at the different time points with the protease inhibitor cocktail.

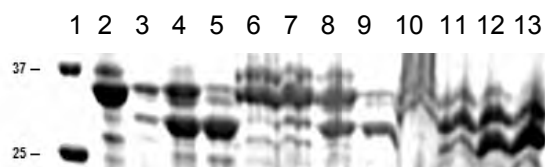


Figure 20: SDS-PAGE showing the degradation pattern of Endo T incubated at 21 °C in 100 mM citric acid buffer pH 3. Lane 2 to 5 contains Endo T without protease inhibitors added, lane 6 to 9 Endo T with protease inhibitors added, lane 10 to 13 Endo T with protease inhibitors and PEG added.

Figure 20 shows a better degradation pattern at pH 3 then figure 16. Samples were taken after 15 minutes, 2 hours and 15 minutes, 9 hours 15 minutes and 1 day and 15 minutes. In all three conditions there is complete degradation of Endo T at pH 3 after only one day. The degradation is slower with protease inhibitors added, but still occurs. This shows us that the inhibitors are not stable or not working at this pH. The quality of the gel was not good and PEG also influences the quality of the bands on SDS-PAGE. This experiment was a control since, proteolytic degradation of Endo T occurred in a pH 3 crystallization condition.

#### 4.3.3 Proteolytic degradation in crystallization

Proteolytic degradation of Endo T was observed in the initial crystal conditions found. These conditions were at pH 2.4 – 3 and as is shown in figure 20 the used protease inhibitors are not stable at this pH. Some crystals and crystal drops were loaded on SDS-PAGE gel to check if Endo T was degraded, this is shown in figure 21.

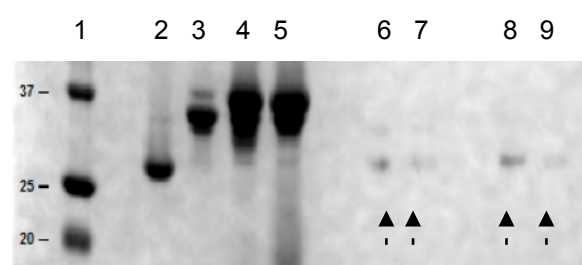


Figure 21: SDS-PAGE showing the protein status in the crystals (Lane 6 and 7) and in the crystal drops (Lane 8 and 9) at pH 3, 100mM citric acid in 5-9% PEG 3350. Lane 1 contains the ladder, lane 2 the active Endo T, lanes 3 to 5 inactive Endo T protein samples Q1F2, Q1F1 and the non-purified Endo T protein sample respectively. Both in the crystals and the crystal drops of Endo T are proteolytic cleaved to their short and active form.

To see if the proteolytic cleavage of Endo T is due to the crystal condition or the protein sample, the protein samples were loaded on gel to see if they were processed. The protein samples with inhibitors were stored for more than 3 months in the fridge. Figure 22 shows us that all these samples are still very similar to the samples taken just after the protein purification.

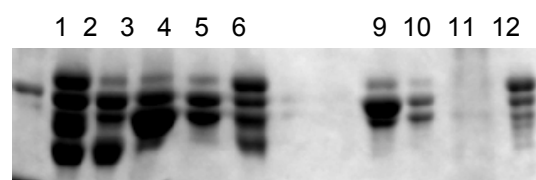


Figure 22: SDS-PAGE of the Endo T protein crystallization samples. With Q2F1 in lane 1; Q2F2; Q1F2; Q1F1 and Q2F1 in lanes 3 to 6. Lane 9 and 10 contain Q1F1 and Q1F2, which were stored in the freezer, finally lane 11 and 12 contain the Endo T (D129A) mutant undiluted and 10 times diluted.

The Endo T (D129A) mutant was loaded to check if this variant of the protein was processed. As shown in lane 12 on figure 22 the Endo T mutant is mainly unprocessed. The proteolytic degradation pattern of both the wild type and the mutant Endo T were then compared. These results are not shown. Due to an error in dilution, the mutant Endo T was not visible on SDS-PAGE. Therefore we cannot compare the degradation pattern of the wild type and the mutant Endo T.

#### 4.3.4 Proteolytic activity on Cel7A.

In a previous paper an acid protease from *T. reesei*, was found to cleave Cel7A and Cel7B into their active form (Eneyskaya, et al. 1999). Since Endo T loses its C-terminus in acid conditions, Endo T was tested to see if it had a protease activity. The acid protease previously discovered has the same molecular mass as Endo T, for these reasons we thought Endo T potentially could possess protease activity in addition to its ENGases activity.

Both the wild type and mutant Endo T were added in different ratio's (1/1, 1/10, 1/20, 1/40, 1/100 and 1/400) to purified intact *T. reesei* Cel7A. All conditions were incubated for 2 hours in 0.1 M pH 3 citric acid buffer. Both the unprocessed and processed Cel7A, by papain cleavage, were added as controls. Cel7A was also incubated without Endo T, as a control for the low pH. These results are not shown. The D129A mutant Endo T was not visible on SDS-PAGE because of a dilution error. There was also no visible protease activity after 2 hours incubation in 0.1 M pH 3 citric acid buffer for all the different ratios.

#### 4.4 Protein crystallization, data collection and structure refinement

Crystallization conditions for Endo T are shown in table 3. These are the crystallization conditions, on which data were collected. The different data collections are shown in table 4. The final statistics for the refined Endo T models are shown in table 5. The final table 6, gives the Ramachandran plot for both new Endo T structures obtained.

Table 3: Different crystallization conditions on which data was collected.

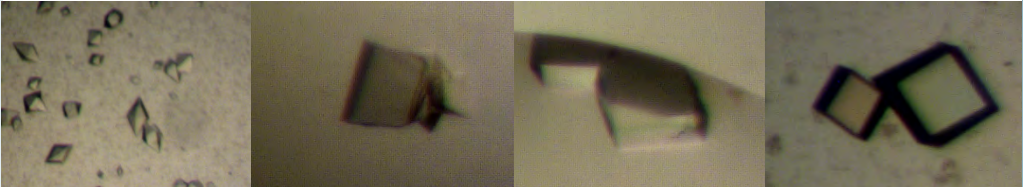
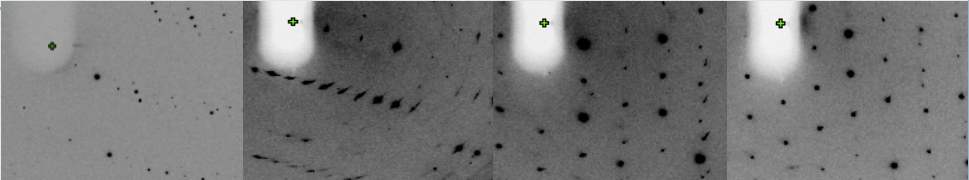
Condition	pH 3	pH 6.5 (a)	pH 6.5 (b)	pH 7.5
Buffer	citric acid	Bis-Tris	Bis-Tris	Hepes
	0.1 M	0.1M	0.1 M	0.1 M
Precipitant	PEG 3350	PEG 3350	PEG 3350	NaPO <sub>4</sub> , PoPO <sub>4</sub>
	5-9%	20 %	20 %	0.8M
Protein	MonoQ	MonoQ	MonoQ	MonoQ
	6-10 mg/ml	8 mg/ml	8 mg/ml	8 mg/ml
Ion		CaCl <sub>2</sub> , CdCl <sub>2</sub> , CoCl <sub>2</sub>	ZnCl <sub>2</sub>	
		0.02M	0.02M	
Time to appear	3 days	1 month	1 month	1 month
Crystal				

Table 4: Data collection on the different crystals shown in table 3.

Condition	pH 3	pH 6.5 (a)	pH 6.5 (b)	pH 7.5
Beam line	I911-3	I911-3	I911-3	I911-3
Wavelength (Å)	1.300	1.300	1.300	1.300
No. Of images	140	140	140	140
Oscillation range (°)	1.0°	1.0°	1.0°	1.0°
Space group	P41	P21	P21	P21
Cell parameters (a,b,c)	89.19,89.19,148.12	35.64,64.23,59.22	35.52,64.05,59.25	35.65,64.18,59.30
Resolution range (Å)	35.22-2.40	23.67-1.31	23.58-1.31	23.65-1.40
Average multiplicity	8.4	2.8	2.8	2.9
Completeness (%)	100	98.1	98.1	99.6
R merge (%)	15.2	5.2	6.3	7.2
I/σ(I)	11.9	11.6	8.5	9.0
Diffraction				

Beam line at MAX-lab, Lund, Sweden

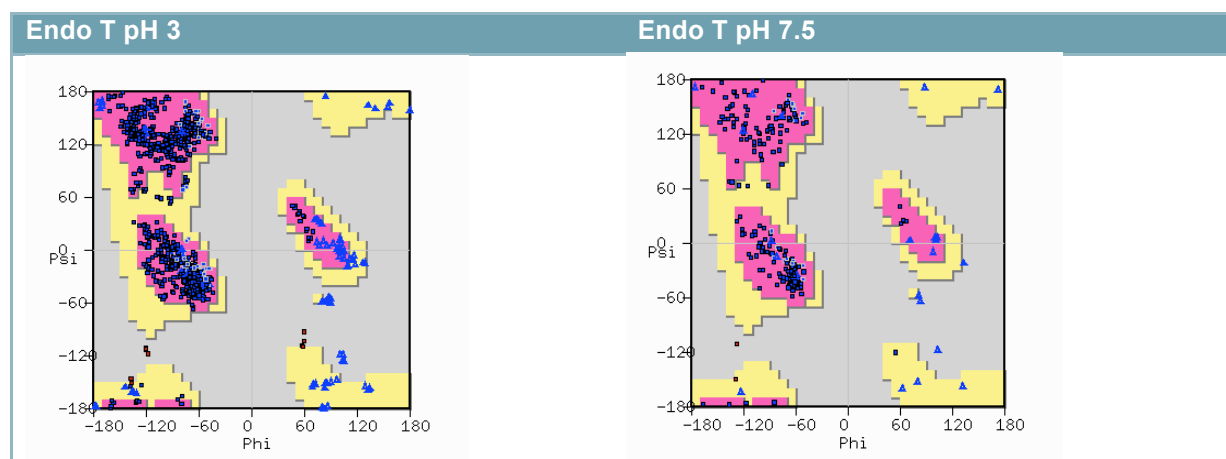
$$R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I - \langle I \rangle|}{\sum_{hkl} \sum_i I}$$

$$R = \frac{\sum |F_o| - |F_c|}{\sum |F_o|}$$

Table 5: Final refinement and model statistics.

	Endo T pH 3	Endo T pH 7.5
Resolution used in refinement (Å)	146.96 - 2.20	23.65 - 1.40
Reflections in: working & test set	54497 - 4029	48821 - 3639
$R^d$ & $R_{\text{free}}$ factor (%)	19.94 – 24.62	13.91 - 16.99
Protein molecules in Asymmetric unit	4	1
Residues in protein	291	287
Protein atoms	10025	2592
Waters	829	272
N-glycosylation (NAG)	8	2
Residues with double conformations	22	14
Average atomic B-factor (Å <sup>2</sup> ):	18.40	14.85
RMSD bond lengths from ideal (Å)	0.0066	0.0085
RMSD bond angles from ideal (°)	1.1956	1.3161
Ramachandran plot statistics (%)		
Preferred regions	95.81	95.56
Allowed regions	3.30	2.67
Outliers	0.89	0.76
Correlation coefficient Fo-Fc (%)	92.9	97.5
Correlation coefficient Fo-Fc free (%)	89.4	96.3

Table 6: Ramachandran plot (coot) for the final refined Endo T structures.



All four crystallization conditions were solved; the pH 3 and pH 7.5 conditions were refined with their final refinement parameters in table 5. All structures were completely similar to previously solved structures, except the structure from the pH 3 condition. In this condition the N-terminus is both longer and in a different conformations in comparison to the three structures. The same protein and protein concentration was used in all 4 conditions. There was no difference between the structures at the C-terminus.



## 4.5 Structural comparison

### 4.5.1 Comparison to known proteins

A structural comparison between the active Endo T and all structural characterized endo-N-acetyl- $\beta$ -D-glucosaminidases (Endo F1, F3 and Endo H) has been carried out in a previous MSc study (Digre 2010). In this study the differences in the active site of Endo T (*T. reesei*), Endo H (*S. plicatus*) and Endo F3 (*E. meningoseptica*) was described. The comparison of the overall structure of Endo T with other ENGases from GH family 18 has been made more recently. This comparison is shown in appendix 4. It was previously mentioned in this thesis that Endo T potentially also could have a protease activity in addition to its ENGases activity. Since Endo T loses its C-terminus very fast at acidic conditions. An acid protease from *T. reesei* with the same molecular mass as Endo T was described in a previous article (Eneyskaya, et al. 1999) Endo T was tested to have a protease activity and screened for a protease active site in the structural comparison shown in appendix 4.

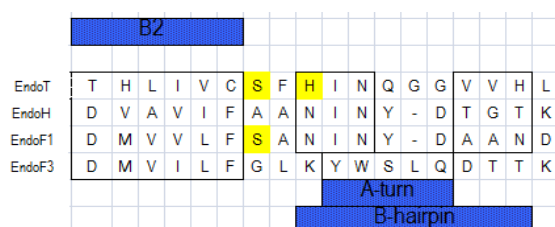


Figure 23: The region in the protein sequence with the possible catalytic protease site in the structurally characterized ENGases. Potential catalytic amino acids in the sequence are colored yellow.

Since the structure of the active Endo T was solved, it was scanned for different protease active site motifs. One potential protease active site was site found, which has the sequence catalytic triade for a serine protease. The distances in the structure are not optimal, but proteins are known to be flexible. The serine and histidine residue of this potential catalytic triade are colored yellow in figure 23. Two possible acid protease active sites can also be found in the structure. It is much harder

to predict the existence and position of an acid protease active site since only two aspartic acid residues are needed to for the active site of an acid protease.

### 4.5.2 Comparison to unknown proteins

A BLAST search with the Endo T sequence (using the BLAST server at Uniprot) against eukaryotic similar enzymes was carried out to compare Endo T to non-structural characterized homologous proteins. The Blast results of the BLAST search are shown in appendix 5. After the BLAST search, the identified sequences that had the highest similarity in amino acid sequence and function (ENGases) were selected and aligned. A genealogic tree and amino acid sequence alignment were generated from this. The genealogic tree is shown in appendix 6, a part of the alignment is shown in appendix 7. This part contains both the ENGase active site and the possible serine protease active site. In the alignment only the six first sequences contain both the serine and histidine residue in the potential protease active site of these proteins, this is shown in Figure 24. Endo FV (*F. velutipes*) was also included in the alignment, but lacks both the serine and histidine residue of the potential serine protease active site, while most of the proteins have the histidine residue. The other proteins, which were used in the sequence alignment are shown in appendices 5 to 7.

THLIVCSLHVNEN-GQIHLND	112	E9EFP8_METAQ
THLIVCSLHVNEN-GQIHLND	111	E9F659_METAR
THLIVCSFHINQG-GVVHLND	114	C4RA89_TRIRE
THLYVCSLHVNYD-GEIHLND	130	A4RJB3_MAGO7
THLIICSLHINKD-GDIHLND	143	D1ZGL1_SORMK
THLIICSIHMHQN-GHLHLND	163	B2AES6_FODAN

Figure 24: BLAST results, which could, just as Endo T have a possible protease active site. Serine and threonine residues are shown in green.

The proteins, which have both the ENGase active site and the possible serine protease active site, come from *Metarhizium acridum* (ENGase, E9EFP8), *Metarhizium robertsii* (ENGase, E9F659), *Trichoderma reesei*

(ENGase, C4RA89), *Magnaporthe oryzae* (unknown function, A4RJB3), *Sordaria macrospora* (unknown function, D1ZGL1) and *Podospira anserine* (unknown function, B2AES6). All of these organisms are filamentous fungi. The protein from *Magnaporthe oryzae* has a threonine residue instead of a serine residue.

#### 4.6 Possible protease active site

Since the structure of the active Endo T was already solved, it was checked for a possible protease active site. There were no free cysteine residues, which would exclude a potential cysteine proteases. One possible serine protease site was found, close to the ENGase active site. Two possible acid protease sites were also found, but after checking the structure one was excluded. Since the proteolysis occurs without any metals in the protein sample we can also exclude a possible metallo protease activity. Other protease activities are more specific and too difficult to screen in an alignment or a known structure. There is though no direct proof yet that Endo T actually possesses both ENGase and protease activity.

##### 4.6.1 Serine protease

The possible serine protease (SP) active site consists of Ser 44, His 46 and possibly Asp 57 and Gly 87. This possible SP (colored blue in figure 25) site is located right next to the ENGase active site (colored red in figure 25). In between these two possible activity sites of Endo T is the loop going from  $\beta$ -sheet to  $\alpha$ -helix 3, which forms some sort of barrier between the ENGase active site of Endo T and the potential serine protease active site of the enzyme.

In the ligand structure of Endo T, there were interactions between His 46 and the branching Man. The suggested Asp also interacted with this Man. There were no interactions with other residues in the protein. Figure 26 is the catalytic site of the protease trypsin a serine protease from pig (PDB code 1AN1). There are

differences in distance and conformation between the amino acids building up the catalytic site of the trypsin protease and the amino acids building up the possible serine protease active site of Endo T. This difference between the active site of the two enzymes might be due to crystal packing interactions causing changes in the position of the potential catalytic amino acids of the two compared structures.

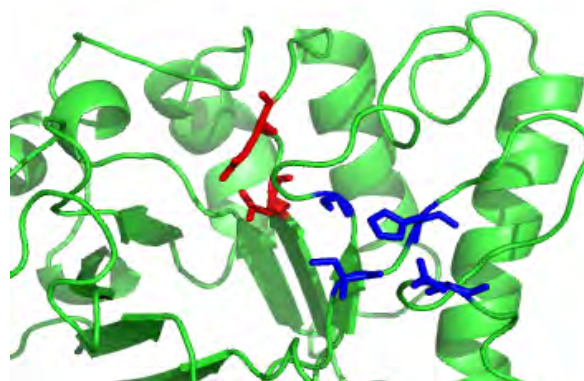


Figure 25: The possible serine protease active site (blue) positioned next to the ENGase active site (red) of Endo T. The cartoon representation of Endo T was made from the Endo T structure with a ligand bound in the enzyme at pH 3.

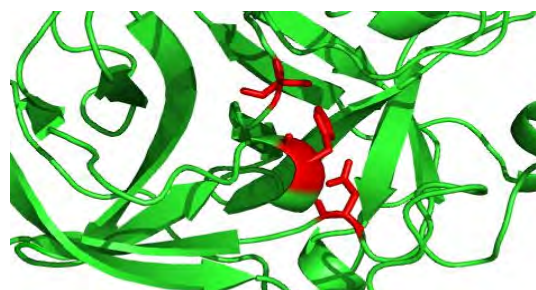


Figure 26: Cartoon representation of the pig trypsin protease catalytic site with the catalytic residues H56, D102 and S196 colored in red.

##### 4.6.2 Acid protease

Two aspartic acids are needed in the active center of an acid protease (AP). After going through the sequence alignment and structure of Endo T, only one potential active site had two Aspartate residues in a conformation where the distances between these two residues could build up an AP active site. The two amino acids that potentially could build up an AP active site are Asp 202 and Asp 205. Figure 27 shows Endo T in green with the

previously described SP active site in blue, the ENGase active site in red and the AP active site in orange compared to the structure of Endo F3 in yellow. The Asp residues are not conserved in the known structures from the GH 18 family, nor in the unknown proteins identified in the BLAST search.

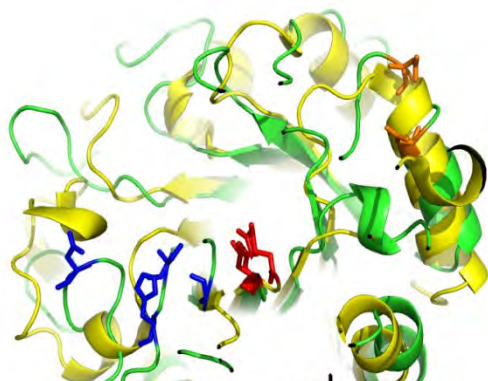


Figure 27: Cartoon representation of Endo T superimposed on Endo F3 in yellow, showing the possible acid protease active site in orange with Asp 202 and Asp 205 in  $\alpha$ -helix 6 of Endo T. The ENGase active residues are shown in red, the previously described serine protease residues in blue.

## 5 Discussion

### 5.1 Biochemical characterization

#### 5.1.1 C-terminal stability

The proteolytic cleavage of the C-terminus is both influenced by temperature and pH. With the addition of a protease inhibitor cocktail we are able to inhibit the proteolytic cleavage, except at very low pH. The bachelor student Jason Blomme did a more extent biochemical characterization (Blomme 2011). He showed that at pH 5 and pH 8 Endo T stays intact for more than 10 days, when the protease inhibitor cocktail is added. This control failed in the first experiments which were carried out. The protein samples were stored in the freezer, where the protease inhibitor cocktail is not stable. Therefore the protease inhibitor cocktail should be added fresh, before each incubation. At pH 5 the proteolysis is enhanced when the intact Endo T is stored at a higher temperature.

When we lower the pH at 21°C the proteolysis is also enhanced. Especially at pH 3 and lower, here complete degradation happens within a day and the protease inhibitor cocktail is not working. The inhibitor might not be stable at this pH or the protease activity is enhanced in such extent that the concentration of the inhibitor is insufficient. This posed big problems for the crystallization experiments, since the only crystallization condition of the active Endo T, which gave crystals for the intact Endo T was at pH 3 and lower. The protein in this crystal was therefore the processed Endo T and not the intact Endo T. This was verified on SDS-PAGE. This pushed us to screen for more crystallization conditions especially at higher pH. Three different crystallization conditions were found, two at pH 6.5 and one at pH 7.5. The actual pH was checked with a pH-strip and was lower than pH 5. The so called higher pH crystallization condition was also the processed Endo T. This was confirmed by SDS-PAGE and the crystals had the same shape as the previous crystals from the active Endo T.

#### 5.1.2 Deglycosylation

The intact Endo T was incubated at 37 °C with protease inhibitors to let the active fraction of the unpurified Endo T sample deglycosylate the intact fraction. Deglycosylated proteins are easier to crystallize, a second reason for performing this experiment, was to see if the deglycosylation and proteolysis are linked or if they take place separately. We can conclude that deglycosylation and proteolysis are linked in fact. After the deglycosylation we had two bands on SDS-PAGE. If they would not be linked, we would only expect one, the completely deglycosylated Endo T with his C-terminus. Since two bands are detected, it seems that only one of the two glycosylation sites was deglycosylated. This was also confirmed by the experiments, which Jason Blomme carried out (Blomme 2011). He suggested that Asn 240, which is close to the cleavage site of the C-terminus, is not as accessible as the Asn 70.



## 5.2 Protease activity

Endo T gets very fast processed, especially at very low pH. The proteolytic cleavage of proteins has been described both in *T. reesei* and *P. pastoris*. We knew this could pose huge problems, even after purification. Therefore it could be that Endo T could have a protease activity. In a previous article from Eneyskaya *et al.* an acid protease from *T. reesei* was described (Eneyskaya, *et al.* 1999). The protein they describe has the same molecular weight as Endo T and behaves in the same way, as problems we encounter when we try to keep Endo T intact. Their protease activity increased massively when the pH was lower than pH 3. They were able to inhibit the protease activity at this low pH, but only when equimolar quantities of pepstatin were added to the protease. Cel7A was one of the substrates Eneyskaya *et al.* tried for this acid protease. After incubating Endo T with Cel7A for two hours at different enzyme/substrate ratios, no proteolytic cleavage was detected. One possible serine protease active site was found when the structure was screened.

## 5.3 Endo T structure

We tried to crystallize the intact Endo T, but we were not able to suppress the proteolysis. Several structures were solved at different pH's. Two models were refined to compare the structure of Endo T at low and more neutral pH.



Figure 28: Comparison of the N-terminus for the different Endo T structures at different pH. In green the Endo T structure at pH 7.5 and in blue the Endo T structure at pH 3.

### 5.3.1 N-terminus

There were no big differences in the structure, except at the N-terminus. Endo T was expressed in *P. pastoris* and since we used a construct, four more amino acids (EAEA) are found at the N-terminus. As is shown in figure 30, the orientation of the N-terminus changed. The longer N-terminus goes towards the ENGase active site of another Endo T molecule in the crystal structure. This did not change in the structure of Endo T at pH 7.5, where the four extra amino acids were not detected in the electron density map. The crystal at pH 3 had a different form (tetrahedral) then the crystal in the other crystallization conditions (cubic). Therefore it is possible that due to the pH and maybe enhanced protease activity, the N-terminus is folding to the molecule next to it. We don't know if the structure at pH 7.5 lacked these four amino acids or that the N-terminus was too floppy to be seen in the electron density map.

### 5.3.2 C-terminus

Since we were not able to control the proteolysis, we were not able to solve the structure of the intact Endo T. Previously was shown that Endo T becomes active when it is both deglycosylated and proteolytically cleaved. Where the proteolytic cleavage is the main drive for activation. This activation is strongly pH dependent. There are two hypotheses why the intact Endo T is inactive. The 47 amino acid long C-terminus could fold back to the top of the  $(\beta/\alpha)_8$  TIM barrel. A second hypothesis is that the C-terminus induces conformational changes in the helices that disturb the ENGase active site at the top of the  $(\beta/\alpha)_8$  TIM barrel. When we look at figure 31 the C-terminus is colored dark orange and could fold back towards the active site. There is more room in between the  $\alpha$ -helices at the C-terminus than there is between the other  $\alpha$ -helices. Whether the C-terminus would fold back between the dark orange and shorter helix or between the shorter and light orange helix is not clear, both options could be possible. In figure 32 we can see that there is direct access to the active

site. When we even look closer at figure 33, we can imagine that when the C-terminus folds back, the substrate might not be able to reach the active site, since it might be hindered by the C-terminus. The second hypothesis is that the C-terminus could induce conformational changes in one of the three helices shown on figures 31 to 33. The short middle helix, has a very long loop at the top. If the C-terminus would change the conformation of this middle helix, the loop could have a very big influence on the active site. Whether it would change the conformation of Asp 129 and Glu 131 or whether it would block the substrate binding is again a pure guess. The right, light orange helix in figure 31 is directly connected to the possible serine protease site. Conformational changes in this helix could alter the conformation of the Ser and His forming a serine protease active site. The helix to the right of the light orange helix in figure 31 is not parallel with the other helices. This helix is turned over approximately 45 °.

### 5.1 Concluding remarks

We can conclude that due to the fact that we could not keep the proteolytic cleavage under control, the structure of the intact Endo T was not solved. Few hypotheses could be made on how the intact Endo T might be activated, but there is no new direct proof supporting these hypotheses. The C-terminus is stable at pH 5 for 6 days, without any inhibitors. At pH 3 it gets processed within 2 hours and is completely processed within the day. This poses a perpetual problem to solve the structure. One hypothesis, is that Endo T could have protease activity. Therefore we tested Endo T for protease activity and screened its structure for a possible protease active site. A possible serine protease active site has been found close to the ENGase active site. The activity of this site might be influenced by the C-terminus itself, since it is so close to the ENGase active site. This and a previous article describing an acid protease from *T. reesei*, which has the same molecular weight as Endo T, suggested us that Endo T could have a protease activity.

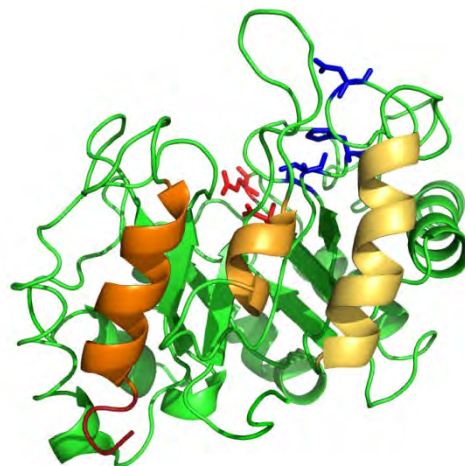


Figure 29: Side view of the short C-terminus. The ENGase active residues are colored red, the possible SP residues are colored blue.

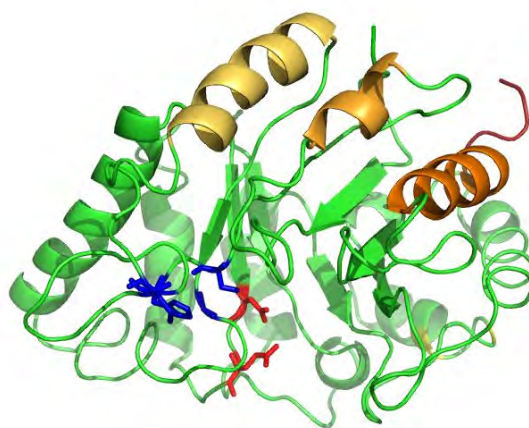


Figure 30: Side top view of the short C-terminus. The ENGase active residues are colored red, the possible SP residues are colored blue.

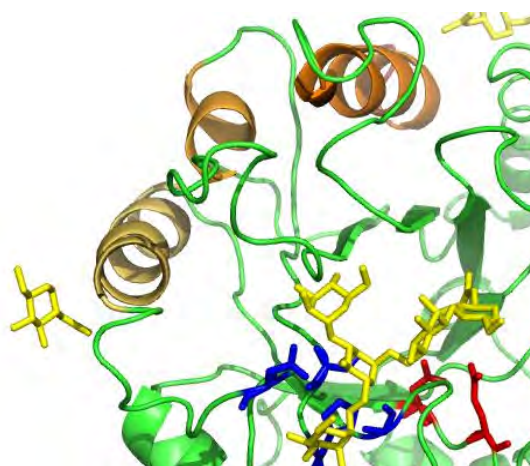


Figure 31: Top view of the short C-terminus. The ENGase active residues are colored red, the possible SP residues are colored blue and the ligand NAG-Man<sub>4</sub> is in yellow.

The protease activity should be tested at lower pH and longer incubation time to determine if Endo T has a protease activity. The specific protease activity should be determined, new protease inhibitors or higher concentration of inhibitors should be tested to inhibit the protease activity. This might control the proteolysis and help to solve the structure of the intact Endo T. Specific mutations could then be made in the suggested protease active sites when the specific protease activity has been found and to determine if the proteolytic cleavage of the intact Endo T is due to itself or another protease.

## **6 Acknowledgments**

Both my supervisors Dr. Mats Sandgren and Dr. Ingeborg Stals gave me the opportunity to perform this interesting research in Uppsala, Sweden. I couldn't have done a fraction what I did and learned without their help and support. I thank Dr. Mats Sandgren specifically for his help with my accommodation, staying, and the scientific and personal problems we discussed in Sweden. I thank Dr. Ingeborg Stals for her help in understanding of the Endo T protein and the scientific discussions. I thank the entire group of molecular biology at the Swedish University of Agricultural Sciences for their help and support and the many coffee brakes we had together. Saeid Karkhehabadi for his help with the purification and physical characteristics of X-ray crystallography. Nils Mikkelsen for his eternal patience and time he spent answering all my questions in solving the structure of the Endo T protein. Roland Berghdahl for all the moments we spent together during experiments.

## 7 References

- Berg, J. M., J. L. Tymoczko, and L. Stryer. *Biochemistry lecture notebook*. W H Freeman & Co, 2006.
- Bergfors, Terese M. *Protein crystallization second edition*. La Jolla: International University Line, 2009.
- . *Protein Crystallization: Techniques, strategies and tips*. La Jolla: International University Line, 1999.
- Biochemica, Boehringer Mannheim. "DIG Glycan differentiation kit." 1996.
- Bio-rad. "Bio-rad protein assay."
- Blomme, Jason. *Deglycosylatie, proteolyse en activatie van het endo-N-acetylglucosaminidase uit T. reesei*. Gent: Hogeschool Gent, 2011.
- Boel, Jonathan. "Structuur-functie onderzoek van een endo-N-acetyl-beta-D-glucosaminidase." Hogeschool Gent, Gent, 2010.
- Branden, Carl, and John Tooze. "Alpha/Beta structures." Chap. 4 in *Introduction to Protein Structure*, by Carl Branden and John Tooze. New-York: Garland-Publishing, 1999.
- Burnette, Neal W. "'Western Blotting': Electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A." 1981: 195-203.
- Calmels, T.P.G., F. Martin, H. Durand, and G. Tiraby. "Proteolytic events in the processing of secreted proteins in fungi." *Journal of biotechnology*, 1991: 51-66.
- Cereghino, Joan Lin, and James M. Cregg. "Heterologous protein expression in the methotrophic yeast *Pichia pastoris*." *Federation of European Microbiological Societies*, 2000: 45-66.
- Chayen, Naomi E. "Turning protein crystallization from an art into a science." *Current Opinion in Structural Biology*, 2004: 577-583.
- . "Methods for separating nucleation and growth in protein crystallization." *Progress in Biophysics & Molecular Biology*, 2005: 329-337.
- . *Protein crystallization strategies for structural genomics*. La Jolla: International University Line, 2007.
- Cregg, James M., Joan Lin Cereghino, Jianying Shi, and David R. Higgins. "Recombinant protein expression in *Pichia pastoris*." *Molecular biotechnology*, 2000: 23-52.
- Davies, Gideon, and Bernard Henrissat. "Structures and mechanisms of glycosyl hydrolases." *Structure*, 1995: 853-859.
- Deshpande, Nandan, Marc R. Wilkins, Nicolle Packer, and Helena Nevalainen. "Protein glycosylation pathways in filamentous fungi." *Glycobiology*, 2008: 626-637.
- Dienes, Dóra, et al. "Identification of a trypsin-like serine protease from *Trichoderma reesei* QM9414." 2007: 1087-1094.

Digre, Andreas. "Mutational Engineering, Growth Selection, Expression, Purification and Crystallization of Ligand Bound Endo-beta-N-acetylglucosaminidase T (EndoT) from *Hypocrea jecorina*." Swedish university of agricultural sciences, Uppsala, 2010.

Dosanjh, Amarjit. *Why Pichia pastoris ?* 1996. [wwwchem.csustan.edu/chem44x0/SJBR/amarjit.htm](http://wwwchem.csustan.edu/chem44x0/SJBR/amarjit.htm) (accessed 2010 11-November).

Eneyskaya, E. V., A. A. Kulminskaya, A. N. Savel'ev, N. V. Savel'eva, K. A. SHabalin, and K. N. Neustroev. "Acid protease from *Trichoderma reesei*: limited proteolysis of fungal carbohydrases." 1999: 226-231.

*European biofuels TECHNOLOGY PLATFORM*. <http://www.biofuelstp.eu/fuelproduction.html> (accessed 2010 11-November).

Goller, Sabine P., Doris Schoisswoll, Michael Baron, Martine Parriche, and Christian P. Kubicek. "Role of endoproteolytic dibasic proprotein processing in maturation of secretory proteins in *T. reesei*." *Applied and environmental microbiology*, 1998: 3202-3208.

Hagspiel, Karl, Doris Haab, and Christian P. Kubicek. "Protease activity and proteolytic modification of cellulases from *Trichoderma reesei* QM414 selectant." 1989: 61-67.

Hamaguchi, Tasuku, Tsukasa Ito, Yukako Inoue, Tipaporn Limpaseni, Piamsook Pongsawasdi, and Kazuo Ito. "Purification, characterization and molecular cloning of a novel endo-beta-N-acetylglucosaminidase from the basidiomycete, *Flammulina velutipes*." *Glycobiology*, 2010: 420-432.

HAMPTON, RESEARCH CORP. *PEG/lon • PEG/lon 2 • PEG/lon HT*. 2003. [http://hamptonresearch.com/product\\_detail.aspx?cid=1&sid=30&pid=11](http://hamptonresearch.com/product_detail.aspx?cid=1&sid=30&pid=11) (accessed August 20, 2011).

Harman, G. E. *Biological control*. <http://www.nysaes.cornell.edu/ent/biocontrol/pathogens/trichoderma.html> (accessed 2010 11-November).

Henrissat, B., A. Bairoch, and G. Davies. *Glycoside Hydrolase family classification*. AFMB. 1998. <http://www.cazy.org/Glycoside-Hydrolases.html> (accessed 2010 2-November).

—. *Glycoside Hydrolase family classification*. AFMB. 1998. <http://www.cazy.org/GH73.html> (accessed 2010 15-November).

—. *Glycoside Hydrolase family classification*. AFMB. 1998. <http://www.cazy.org/GH85.html> (accessed 2010 15-November).

—. *Glycoside Hydrolase family classification*. AFMB. 1998. <http://www.cazy.org/GH18.html> (accessed 2010 15-November).

International Union of Crystallography. *Collaborative Computational Project No. 4 Software for Macromolecular X-Ray Crystallography*. 2011. <http://www.ccp4.ac.uk/> (accessed August 20, 2011).

Karamanos, Y. "Endo-N-acetyl-β-D-glucosaminidases and their potential substrates : structure/function relationships." *Microbiological research*, 1997: 661-671.

Lattman, Eaten E., and Patrick J. Loll. *Protein crystallography: a concise guide*. Baltimore: The John Hopkins University Press, 2008.

Lehle, L. "Protein glycosylation in yeast." *Antonie van Leeuwenhoek* 61, no. 2 (1992): 133-134.

Maras, Marleen, et al. "Structural characterization of N-linked oligosaccharides from cellobiohydrolase I secreted by the filamentous fungus *Trichoderma reesei* RUTC 30." 1997: 617-625.

Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D, Danchin EG, Grigoriev IV, Harris P, Jackson M, Kubicek CP, Han CS, Ho I, Larrondo LF, de Leon AL, Magnuson JK, Merino S, Misra M, Nelson B, Putnam N, Robbertse B, Salamov AA, Schmoll M, Terry A, Thayer N, Westerholm-Parvinen A, Schoch CL, Yao J, Barbote R, Nelson MA, Detter C, Bruce D, Kuske CR, Xie G, Richardson P, Rokhsar DS, Lucas SM, Rubin EM, Dunn-Coleman N, Ward M, Brettin TS. "Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei*. ." *Nat Biotechnol.* 26 (2008): 553-560.

McDowall, Jennifer. *Trypsin and chymotrypsin, serine protease digestive enzymes*. David Goodsell. [http://www.ebi.ac.uk/interpro/potm/2003\\_5/Page1.htm](http://www.ebi.ac.uk/interpro/potm/2003_5/Page1.htm) (accessed 06 23, 2011).

Moss, G. P. *E.C. 3.4 acting on peptide bonds*. <http://www.chem.qmul.ac.uk/iubmb/enzyme/EC3/4/> (accessed 2010 12-November).

—. *EC 3.2.1 Glycosidases, i.e. enzymes hydrolysing O- and S-glycosyl compounds*. IUPAC-IUBMB. 1992. <http://www.chem.qmul.ac.uk/iubmb/enzyme/EC3/0201a.html#321> (accessed 2010 26-March).

Navarro, Adeline, Ho-Shing Wu, and Shaw S. Wang. "Engineering problems in protein crystalization." *Seperation & Purification Technology*, 2009: 129-137.

Powell, Harry. *A new interface for Mosflm*. 02 13, 2010. <http://www.mrc-lmb.cam.ac.uk/harry/imosflm/ver104/> (accessed August 20, 2011).

QIAGEN. *Sample and assay technologies*. 2003. <http://www.qiagen.com/products/protein/crystallization/corescreensforinitialinvestigations/jcsgplussuite.aspx?ShowInfo=1> (accessed August 20, 2011).

Selinheimo, Emilia, et al. "Production and characterization of a secreted, C-terminally processed tyrosinase from the filamentous fungus *Trichoderma reesei*." 2006: 4322-4335.

Sharma, Ruchika, Meenu Katoch, P. S. Srivastava, and G. N. Qazi. "Approaches for refining heterologous protein production in filamentous fungi." 2009: 2083-2094.

Sinha, Jayanta, Bradley A. Plantz, Mehmet Inan, and Michael Meagher. "Causes of proteolytic degradation of secreted recombinant proteins produced in ethylotropic yeast *Pichia pastoris*: Case study with recombinant ovine interferon-T." *Biotechnology and bioengineering*, 2005: 102-112.

Stals, Ingeborg, et al. "Identification of a gene coding for deglycosylating enzyme in *Hypocrea jecorina*." *Federation of European Microbiological Societies*, 2009.

Stals, Ingeborg, Koen Sandra, Bart Devreese, Jozef Van Beeumen, and Marc Claeysens. "Factors influencing glycosylation of *Trichoderma reesei* cellulases. II: N-glycosylation of Cel7A core protein isolated from different strains." 2004: 725-737.

Stals, Ingeborg, Koen Sandra, Steven Geysens, Roland Contreras, Jozef Van Beeumen, and Marc Claeysens. "Factors influencing glycosylation of *Trichoderma reesei* cellulases. I: Postsecretorial changes of the O- and N-glycosylation pattern of Cel7A." 2004: 713-724.

Stanley, Pamela, Harry Schachter, and Naoyuki Taniguchi. "N-Glycans." In *Essentials of glycobiology*, by A. Varki, R. Cummings and J. Esko. Cold Spring Harbour (NY): Cold Spring Harbour laboratory press, 2009 2nd edition.

Tarentino, Anthony L., Thomas H. Plummer, and Frank Maley. "The release of intact oligosaccharides from specific glycoproteins by endo- $\beta$ -N-acetylglucosaminidase H." *The journal of biological chemistry*, 1974: 818-824.

Taylor, John W., Joey Spatafora, and Mary Berbee. *The Tree of Life Web Project*. 2006. <http://www.tolweb.org/Ascomycota/20521> (accessed 2010 12-November).

Trimble, Robert B., and Anthony L. Tarentine. "Identification of distinct endoglycosidase (Endo) activities in *Flavobacterium meningosepticum*: Endo F1, Endo F2 and Endo F3." *The Journal of Biological Chemistry*, 1991: 1646-1651.

Tsutomu Takayanagia, Atsuo Kimurab, Seiya Chibab and Katsumi Ajisaka. "Novel structures of N-linked high-mannose type oligosaccharides containing  $\alpha$ -d-galactofuranosyl linkages in *Aspergillus niger*  $\alpha$ -d-glucosidase." *Carbohydrate Research* 256, no. 1 (1994): 149-158.

Voet, Donald J., Judith G. Voet., and Charlotte W. Pratt. "Protein structure and folding." Chap. 6 in *Principles of biochemistry*, by Donald J. Voet, Judith G. Voet. and Charlotte W. Pratt. Wiley, 2008.

Voet, Donald, Judith G. Voet, and Charlotte W. Pratt. "Enzymatic catalysis." Chap. 11 in *Principels of Biochemistry*. Wiley, 2006.

Wlodawer, Alexander, Wladek Minor, Zbigniew Dauter, and Mariusz Jaskolski. "Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures." 2007: 1-21.



## 8 Appendix

### 8.1 Appendix 1: Bio-Rad assay results

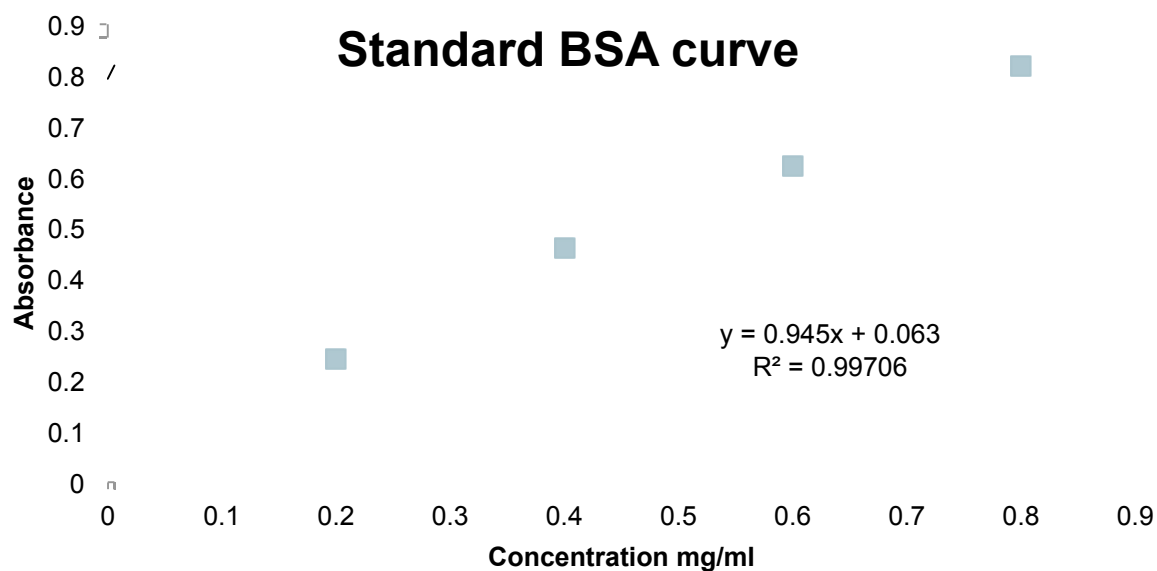


Figure 32: BSA standard curve for total protein concentration.

Table 7: Unpurified Endo T total protein concentrations.

Endo T dilution	Without inhibitors		Endo T	With inhibitors	
	absorbance	concentration		absorbance	concentration
10	0,314	3,143		0,887	8,868
20	0,126	2,529		0,505	10,095
50				0,185	9,259
	average	concentration		average	concentration
		2,836			9,407

### 8.2 Appendix 2: Purified protein concentrations

Table 8: Purified Endo T protein concentrations.

Fraction	Q1F1	Q2F1	Q1F2	Q2F2
Absorbance	19.81	8.56	14.46	5.68
Concentration (mg/ml)	13.76	5.94	10.00	3.94
Final volume	< 500 µl	± 750 µl	± 200 µl	± 750 µl
260/280 value	0.68	0.68	0.71	0.65

QxFy; x stands for the batch, while the y stands for the peak collected.



### 8.3 Appendix 3: Western blot SDS-PAGE

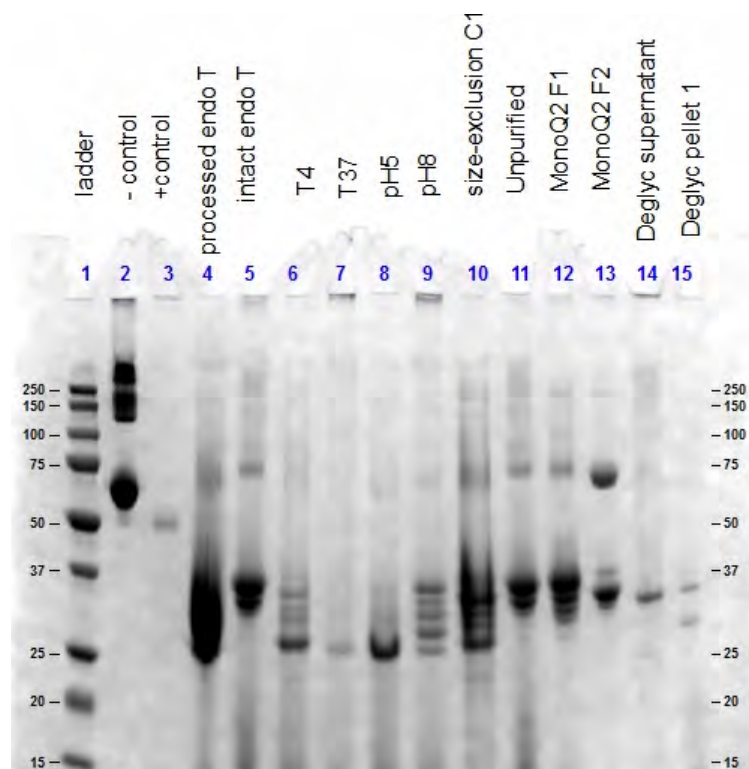


Figure 33: First SDS-PAGE for western blot.

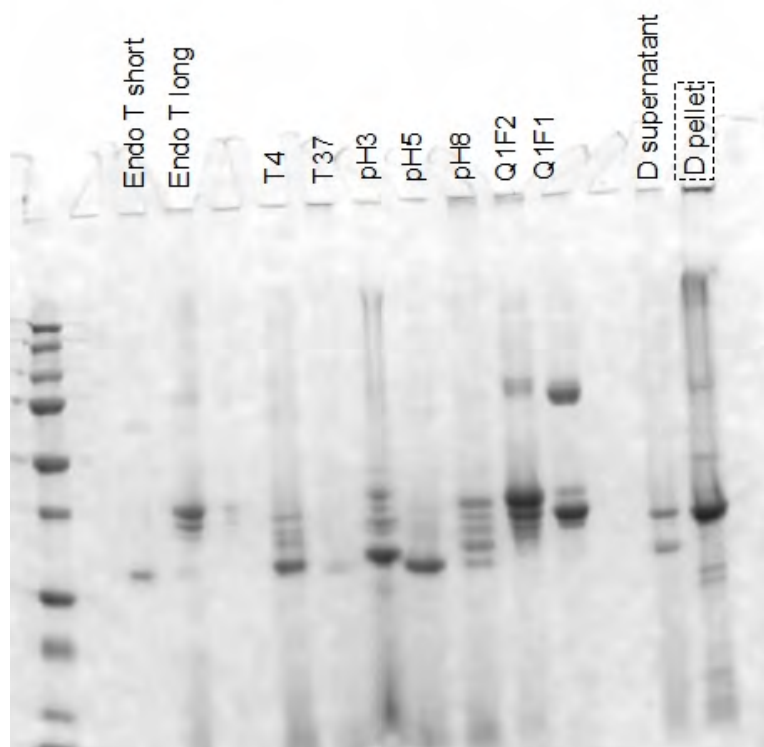


Figure 34: Second SDS-PAGE for western blot.

#### 8.4 Appendix 4: Structural comparison of known endo- $\beta$ -D-glucosaminidases

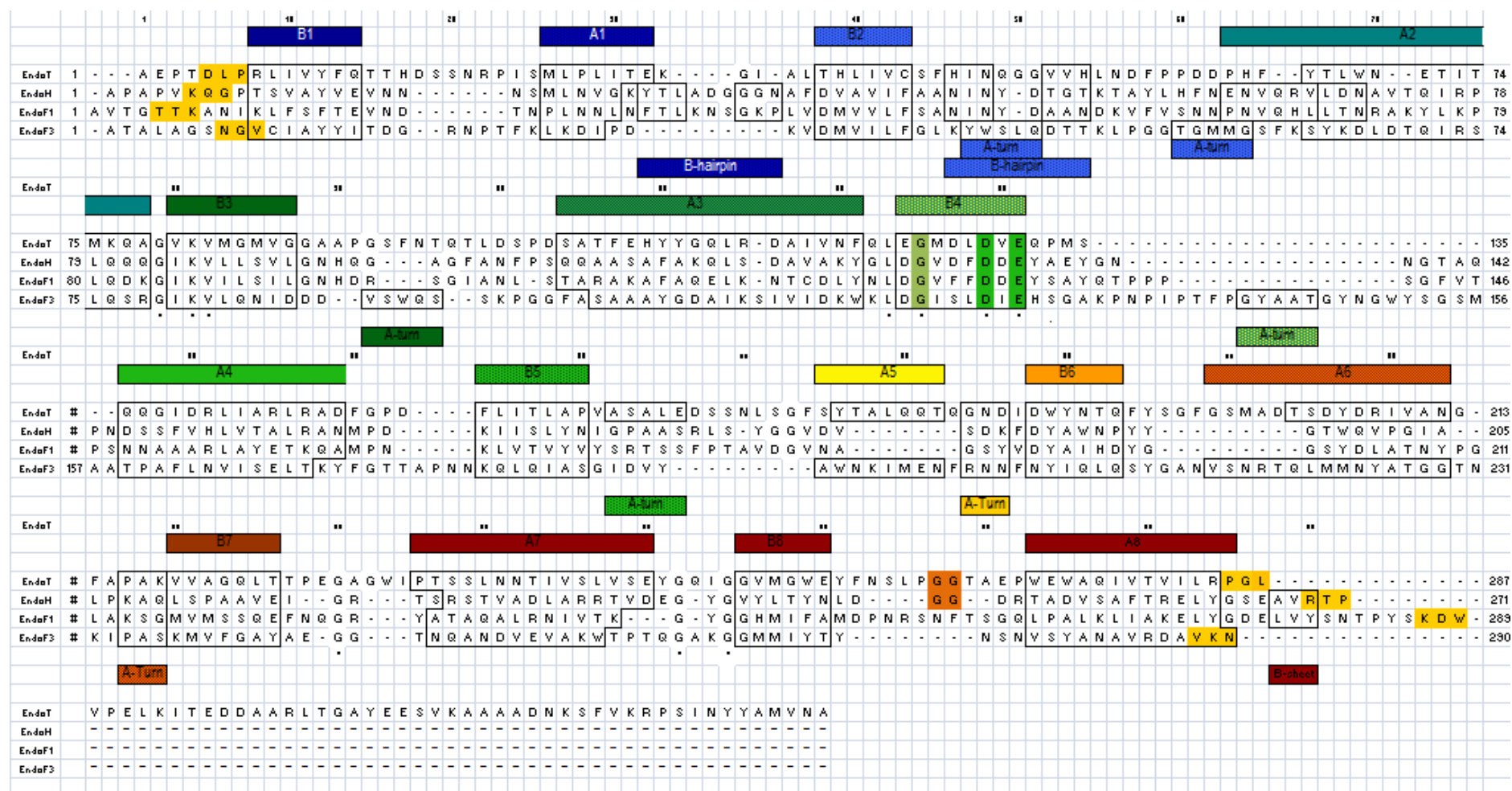


Figure 35: Structural comparison of known endo- $\beta$ -D-glucosaminidases.

## 8.5 Appendix 5: Endo T eukaryotic blast results

Accession	Entry name	Protein names	Organism	Gene names
<a href="#">E9EFP8</a>	E9EFP8_METAQ	Endo-N-acetyl-beta-D-glucosaminidase	Metarhizium acridum (strain CQMa 102)	MAC_08696
<a href="#">E9F659</a>	E9F659_METAR	Endo-N-acetyl-beta-D-glucosaminidase	Metarhizium robertsii (strain ARSEF 23) (Metarhizium anisopliae)	MAA_07758
<a href="#">C4RA89</a>	C4RA89_TRIRE	Endo-N-acetyl-beta-D-glucosaminidase	Trichoderma reesei (Hypocrea jecorina)	endoT
<a href="#">A4RJB3</a>	A4RJB3_MAGO7	Putative uncharacterized protein	Magnaporthe oryzae (strain 70-15 / FGSC 8958) (Rice blast fungus) (Pyricularia oryzae)	MGG_01876
<a href="#">D1ZGL1</a>	D1ZGL1_SORMK	Whole genome shotgun sequence assembly, scaff...	Sordaria macrospora (strain ATCC MYA-333 / DSM 997 / K(L3346) / K-hell)	SMAC_05689
<a href="#">B2AES6</a>	B2AES6_PODAN	Predicted CDS Pa_5_1650	Podospora anserina (strain S / DSM 980 / FGSC 10383) (Pleurage anserina)	
<a href="#">A7E885</a>	A7E885_SCLS1	Putative uncharacterized protein	Sclerotinia sclerotiorum (strain ATCC 18683 / 1980 / Ss-1) (White mold) (Whetzelinia sclerotiorum)	SS1G_01513
<a href="#">A6RYP5</a>	A6RYP5_BOTFB	Putative uncharacterized protein	Botryotinia fuckeliana (strain B05.10) (Noble rot fungus) (Botrytis cinerea)	BC1G_05559
<a href="#">A4R501</a>	A4R501_MAGO7	Putative uncharacterized protein	Magnaporthe oryzae (strain 70-15 / FGSC 8958) (Rice blast fungus) (Pyricularia oryzae)	MGG_04073
<a href="#">E9F229</a>	E9F229_METAR	Alkaline phosphatase	Metarhizium robertsii (strain ARSEF 23) (Metarhizium anisopliae)	MAA_06227
<a href="#">E9EDC1</a>	E9EDC1_METAQ	Alkaline phosphatase	Metarhizium acridum (strain CQMa 102)	MAC_07869
<a href="#">C7YUP6</a>	C7YUP6_NECH7	Glycoside hydrolase family 18	Nectria haematococca (strain 77-13-4 / FGSC 9596 / MPVI) (Fusarium solani subsp. pisi)	NECHADRAFT_49128
<a href="#">Q871X2</a>	Q871X2_NEUCS	Related to chitinase	Neurospora crassa	B7O17.040
<a href="#">F5HFI3</a>	F5HFI3_NEUCR	Putative uncharacterized protein	Neurospora crassa (strain ATCC 24698 / 74-OR23-1A / CBS 708.71 / DSM 1257 / FGSC 987)	NCU01393
<a href="#">D1ZUP9</a>	D1ZUP9_SORMK	Whole genome shotgun sequence assembly, scaff...	Sordaria macrospora (strain ATCC MYA-333 / DSM 997 / K(L3346) / K-hell)	SMAC_12618
<a href="#">B2B6W7</a>	B2B6W7_PODAN	Predicted CDS Pa_2_9140	Podospora anserina (strain S / DSM 980 / FGSC 10383) (Pleurage anserina)	
<a href="#">E3QZ63</a>	E3QZ63_COLGM	Chitinase	Colletotrichum graminicola (strain M1.001 / M2 / FGSC 10212) (Maize anthracnose fungus) (Glomerella graminicola)	GLRG_11295
<a href="#">C9SNX0</a>	C9SNX0_VERA1	Chitinase	Verticillium albo-atrum (strain VaMs.102) (Verticillium wilt)	VDBG_06595
<a href="#">C5FI04</a>	C5FI04_ARTOC	Chitinase 3	Arthroderma otae (strain CBS 113480) (Microsporum canis)	MCYG_01803
<a href="#">E4US03</a>	E4US03_ARTGP	Chitinase 3	Arthroderma gypseum (strain ATCC MYA-4604 / CBS 118893) (Microsporum gypseum)	MGYG_04265
<a href="#">D5GG35</a>	D5GG35_TUBMM	Whole genome shotgun sequence assembly, scaff...	Tuber melanosporum (strain Mel28) (Perigord black truffle)	GSTUM_00001975001
<a href="#">D8QCH6</a>	D8QCH6_SCHCM	Glycoside hydrolase family 18 protein	Schizophyllum commune (strain H4-8 / FGSC 9210) (Split gill fungus)	SCHCODRAFT_236684
<a href="#">D1GA49</a>	D1GA49_FLAVE	Endo-beta-N-acetylglucosaminidase	Flammulina velutipes	endo
<a href="#">B0DHY7</a>	B0DHY7_LACBS	Glycoside hydrolase family 18 protein	Laccaria bicolor (strain S238N-H82) (Bicoloured deceiver) (Laccaria laccata var. bicolor)	LACBIDRAFT_185397

Figure 36: Endo T eukaryotic blast results.

## 8.6 Appendix 6: Genealogic tree from Endo T BLAST results

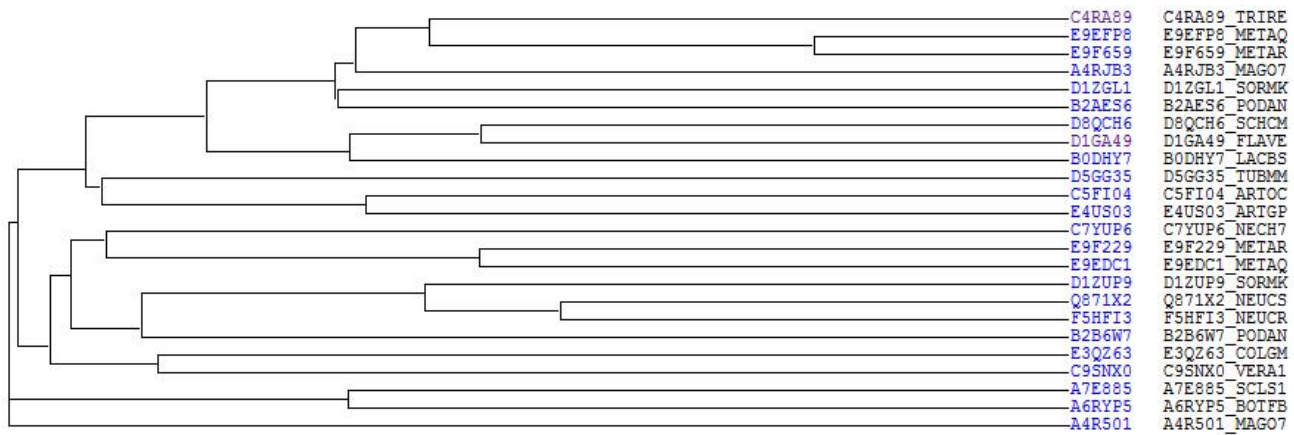


Figure 37: Genealogic tree from Endo T BLAST results.

## 8.7 Appendix 7: Sequence alignment from Endo T BLAST results

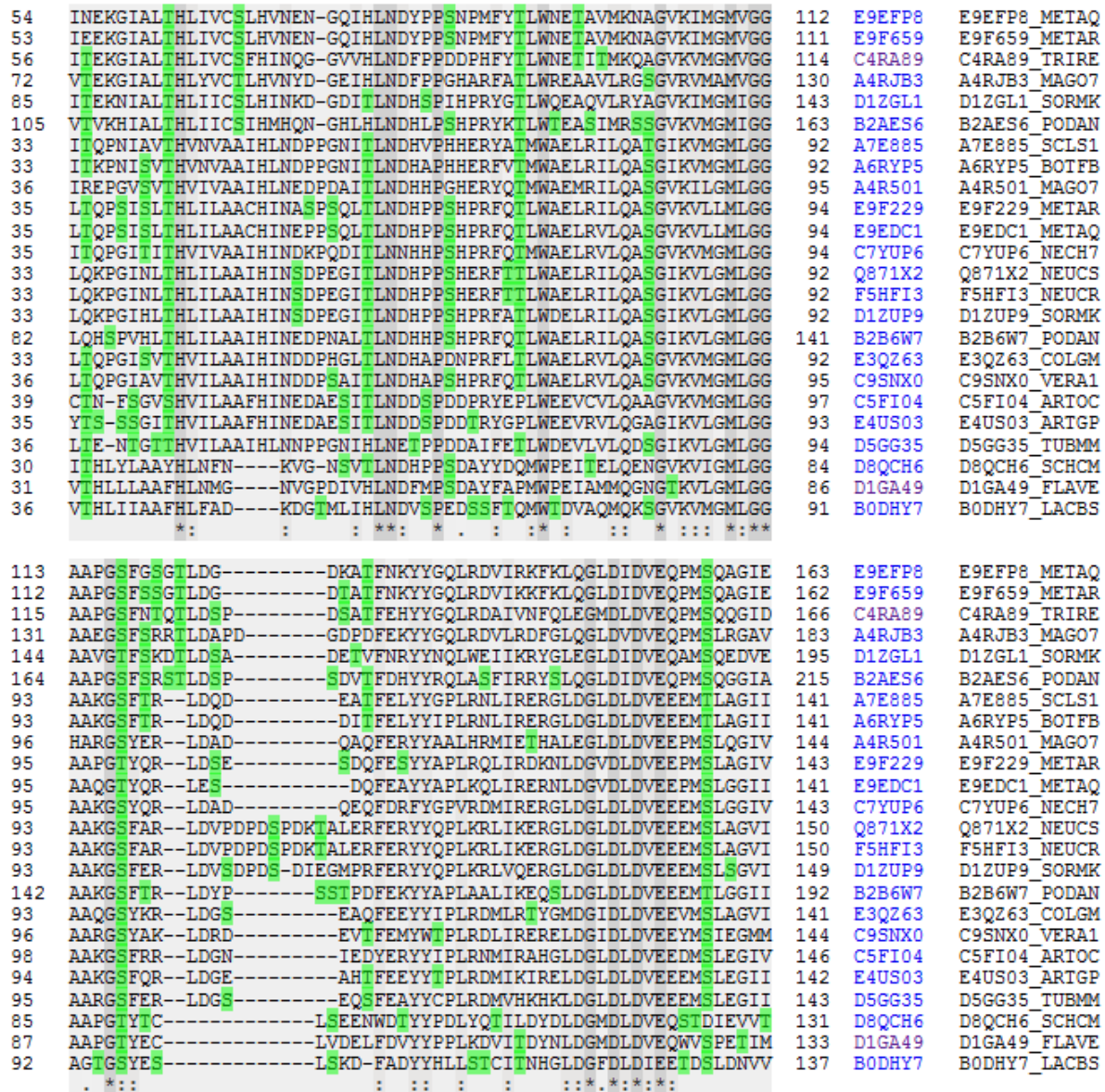


Figure 38: Sequence alignment from Endo T BLAST results.